Mathematical Modeling of Length Regulation in the Type III Secretion System

by

Maulik Nariya

Submitted to the graduate degree program in Department of Physics and Astronomy and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Committee members:

Jack J. Shi (Co-chair) _____

Eric J. Deeds (Co-chair) _____

Chistopher Fischer _____

Philip Bringer _____

Roberto De Guzman _____

Date defended: July 30, 2018

The Dissertation Committee for Maulik Nariya certifies that this is the approved version of the following dissertation:

Mathematical Modeling of Length Regulation in the Type III Secretion System

Jack J. Shi, Co-chair

Eric J. Deeds, Co-chair

Date approved: July 30, 2018

Abstract

Type III secretion secretion (T3SS) system is a protein export pathway that helps bacterial cells construct many structures, like the flagellar hook and the injectisome, that aid in crucial physiological processes such as locomotion and pathogenesis. Both, the flagellar hook and the injectisome, involve long extracellular channels and the length of these channels is highly regulated to allow these structures to perform their intended functions. Numerous experiments have been performed to understand the structural details of this nanomachine during the past decade. Despite the concerted efforts of molecular and structural biologists, several crucial aspects of the assembly of these structure, such as the regulation of the length of the needle and the flagellar hook, remain unclear. There are two leading models for how length control is achieved in the flagellar hook and T3SS needle: the substrate switching model, where the length is controlled by assembly of an inner rod, and the ruler model, in which a molecular ruler controls the length. While there is qualitative experimental evidence to support both models, there is a lack of detailed quantitative characterization of these models that could be used to unambiguously test these mechanisms experimentally. In this work, we used a combination of mathematical and computational techniques to better understand these length control mechanisms.

Based on a set of straightforward assumptions, we constructed a mathematical model for length control based on the timing of substrate switching. Our model made predictions about commonly measured quantities such as the average needle lengths and the variance in lengths. In particular our model predicted for the substrate switching mechanism that the variance scales quadratically with the average length. Our model also predicted the form of the needle length distribution based on this mechanism, and found excellent agreement with available experimental data from *Salmonella typhimurium* with only a single free parameter.

We also constructed a mathematical model of length control based on the ruler mechanism, and found that the predictions of this model are consistent with experimental data not just for the scaling of the average length with the ruler protein length, but also the variance. Interestingly, we found that the ruler mechanism allows for the evolution of needles with large average lengths without the concomitant large increase in variance that occurs in the substrate switching mechanism. In addition to making further predictions that can be tested experimentally, these findings shed new light on the trade-offs that may have lead to the evolution of different length control mechanisms in different bacterial species.

Acknowledgements

First and foremost, I would like to thank my parents, Kalpana and Bhopalsingh Nariya, for continued help and support to pursue my goals in life, this would not have been possible without you. To my dad: a lot of this work is an outcome of values that you cultivated in me. To my mom: I have flourished as an individual through your constant love and devotion.

I would like to thank my advisors, Eric Deeds and Jack Shi. Very few graduate students are fortunate enough to have two advisors for their doctoral work. My dissertation is a unique blend of the expertise of my advisors in their respective fields and it is needless to say that it would not have been possible without their kind guidance. My hope is to live up to their expectations in my future academic endeavors. Eric, thank you for introducing me to the world of systems biology, teaching me the tricks of the trade, and giving me the opportunity to build a career in biology. Jack, thank you for teaching me all the "mathematical wizardry", it is a valuable skill that will stay with me for life. I would also like to thank the remainder of my dissertation committee, Christopher Fischer, Philip Baringer, and Roberto De Guzman, for all their time and guidance. Kyongchul Kong, thank for being my advisor through my initial research interests, I have acquired useful skills from our work together. I would also like to thank the department Chair, Hume Feldman and the Director of Graduate Studies, Gregory Rudnick, not only for providing a platform to pursue interdis-

ciplinary research but also for all the support and encouragement in doing so. And last but not the least, I would like to thank the administrative staff: Kristin Rennells, Teri Leahy, Kim Hubble, Desirée Neyens, Joel Sauerwein, Kayla Wegley, and Debbie Douglass-Metsker (Center for Computational Biology).

All the past and current lab members: Michael Rowland, Ryan Suderman, Joe Greenbaum, Johnny Israeli, Abhishek Mallela, Pushpa Itagi, Anupama Kante, and Shamus Cooley, thank you all for being such amazing colleagues and for the valuable exchange of intellectual expertise. Michael and Ryan, thank you for your help with KaSim, BNGL, Python. Abhishek, thank you for your help with the analytical calculations for the ruler model. A special thanks to Pushpa and Anupama for all the Indian food.

I would like to thank all my friends in Lawrence who have supported me during the course of my time here. Sinan Değer, thank you for being a constant companion in my journey through graduate school. Marianna Simões, thanks for all the Karaoke Nights which helped me blow the steam off. Gopolang Mohlabeng, thank you for continued friendship after being a colleague in particle physics. Kenneth Avocetien, thank you for editing my dissertation. Cosmos Crew: Ramón Alvarado, Trevor Swanson, and Laura Jiménez, thank you all for feeding my curiosities in science, philosophy, music, and art, among other things. Carey Bowen, thank you for making it feel like home in a foreign land.

Contents

List of Figures

List of Tables

Chapter 1

Type III Secretion System: An Overview

1.1 Introduction

An average sized bacterial cell, such as *Escherichia coli* contains over a million proteins [1]. These proteins play important roles in a wide range of cellular functions such as DNA replication, motility, pathogenesis, and chemotaxis, to name a few. In order to execute these functions, the bacterial cells have to transport the proteins to different regions within the cell as well as across their inner, and in case of gram-negative bacteria, the outer membranes. To transport these proteins, bacteria have evolved a wide variety of extremely sophisticated mechanisms, which are referred to as the secretion systems or secretion pathways. Some of these secretion systems consist of complex macromolecular structures that are assembled by bacteria with an extremely high degree of precision. Protein secretion plays an important role in the interaction between the bacteria and their environment and there has been a growing interest in understanding the structure and function of the bacterial secretion systems in the

past two decades. Of the currently known protein export pathways, the type III secretion system (T3SS), which is a major virulence factor present in most gram-negative pathogens, has gained much deserved attention in the past decade. A thorough understanding of such export pathways could lead to a new, more efficacious class of antibiotics. High resolution imaging such as transmission electron microscopy (TEM) has uncovered many intriguing aspects about the structure of this nanomolecular machinery, and significant progress has been made in understanding its assembly mechanisms. However, there are still numerous gaps in our current understanding of T3SS and researchers are trying to gain insights on a fundamental level about this export pathway. The T3SS apparatus has an extremely intricate structural organization, and many components of this machinery have still not been fully characterized. Furthermore, the assembly of these structures is a complex, dynamical process involving numerous interactions on molecular scales, and current experimental techniques, such as cryo-EM, only provide a static snapshot of this assembly. Therefore, it is imperative that we adopt alternate methods to gain further insight and better understand complex protein export pathways. This chapter provides a brief overview of the T3SS, structural details of its components, and its role in assembly of essential structures such as the flagellum and the injectisome. We shall also learn about the most popular length control mechanisms for T3SS and highlight the need for a mathematical description of these proposed mechanisms.

Bacterial secretion systems are classified on the basis of their structure, functions, and substrate specificity. Some of these secretion systems are present in all classes of bacteria and secrete a wide range of proteins, while others are present only in a small subset of bacterial species [2]. The general secretion (Sec) and the twin argenine translocation (Tat) pathways are the most basic of the secretion systems that are present in both gram-negative

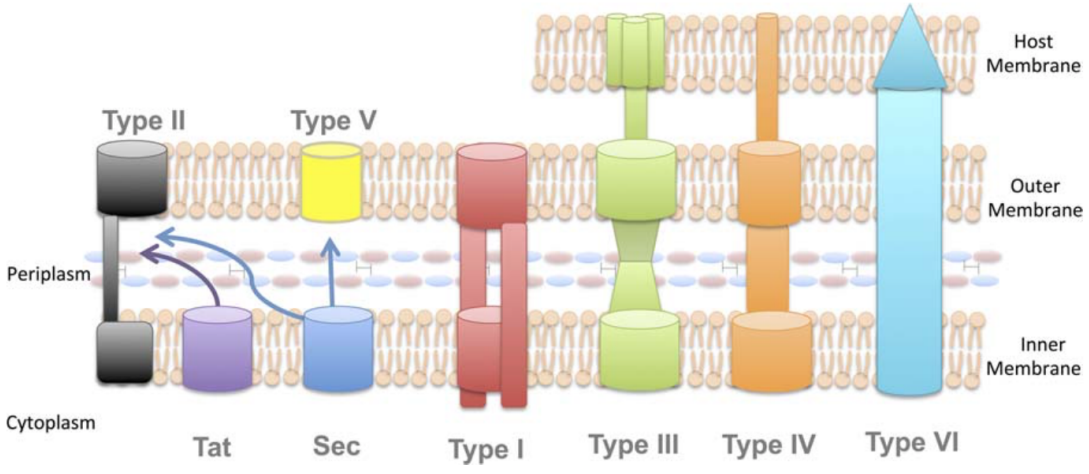


Figure 1.1: Schematic of secretion systems in gram-negative bacteria. Figure source: ref. [2]

as well as gram-positive bacteria. Sec and Tat pathways transport proteins (in unfolded and folded states, respectively) across the inner membrane [3]. Type I–VI secretion systems (T1SS–T6SS) are specific to gram-negative bacteria, which consist of two phospholipid membranes: the inner and outer membranes. Each of these systems is involved in transport of a distinct subset of proteins across the two, or sometimes three, membranes [2]. The T1SSs exports proteins of various sizes and functions across both the inner and outer membranes into the extracellular region as a one-step process. T1SSs have three main structural components: ATP-binding cassette (ABC) transporter proteins, which export small molecules such as antibiotics out of the cell, membrane fusion protein (MFP), which connects the inner and the outer membranes, and outer membrane factor, which generates a pore in the outer membrane [4]. Some bacterial species have several T1SSs, each secreting one or a few proteins [5]. The T2SS apparatus is present only in the outer membrane, and as a result it is capable of transporting only those proteins that are already present in the periplasm (the space between the inner and the outer membranes) into the extracellular region. Originally T2SS was thought to be the terminal branch of Sec pathway, but recent studies show that it is involved in the transfer of proteins secreted through the Sec and the Tat pathways [6, 7].

T3SS is found in most pathogenic gram-negative bacteria. The hallmark of this secretion system is the injectisome, a needle-like structure that is used to transfer effector proteins into the cytoplasm of the host cell during pathogenesis [8]. Along with assembly of the injectisome, the T3SS export apparatus is also involved in assembly of the flagellum, and it is widely accepted that the injectisome and the flagellum share a common ancestry [9, 10]. T4SS is evolutionarily related to the bacterial DNA conjugation systems, and secretes single proteins, protein-protein complexes and DNA-protein complexes [11]. Like T3SS, T4SS has an extracelluar pilus which allows transfer of proteins across a host cell membrane [2]. Structural details of T4SS are only available for one class, known as type IVA. T4SS is most well-studied in *Argobacterium tumeficans* consisting of 12 proteins, VirB1–VirB11 [12]. T5SS is unique in that it does not possess a dedicated secretion apparatus, instead the proteins secrete themselves. The proteins have their own β-barrel domain which inserts itself into the outer membrane forming a transport channel through which the remainder of the protein (or a separate protein) secretes itself [13, 14]. T5SSs can be classified into three types depending on the number of proteins involved in the secretion process: the autotransporter secretion, two-partner secretion, and chaperone-usher secretion [14]. T6SS is the most recently discovered and the least understood of all the secretion systems. A T6SS transports proteins into a variety of cells, including eukaryotic cells, but primarily into other bacterial cells, and is thought to play a role in bacterial communication [15]. A T6SS consists of an inverted phage tail-like structure that is involved in protein transport across the bacterial cell membrane into the target cells [15, 16]. Figure (1.1) shows a schematic representation of the secretion systems present in gram-negative bacteria.

Gram-positive bacteria have only one phospholipid membrane and are surrounded by a very thick cell wall. Because of the structural differences in these cells, gram-positive bacteria em-

ploy different mechanisms for protein secretion. In order to facilitate protein secretion into the external environment, the proteins must be embedded in the cell wall. To achieve this function, most gram-positive bacteria produce enzymes know as sortases, which covalently bind the proteins to the cell wall [17, 18]. In addition to the cell wall, some gram-positive bacteria contain a very dense, hydrophobic lipid layer, known as the mycomembrane. Not only does the mycomembrane serve as a barrier and protect the bacteria from the external environment, but it also makes it extremely difficult for the bacteria to transport substrates across this membrane. These bacteria utilize a specialized mechanism called the type VII secretion system (T7SS), also known as the ESX system [19, 20]. The T7SS consists of five core proteins that form the ESX-5 complex, which is present in the inner membrane. In addition, the T7SS is believed to have a mycomembrane transport channel, however this channel has not yet been observed experimentally [21]. Table (1.1) summarizes the known bacterial secretion systems.

The T3SS is an essential virulence factor that is present in many gram-negative pathogens that are responsible for severe diseases such as typhoid fever (*Salmonella typhimurium*), gastroenteritis (*Shigella flexneri*), plague (*Yersinia pestis*), and infantile bacterial diarrhea (enteropathogenic *Escherichia coli* or EPEC) [22]. Intracellular pathogens such as *Salmonella* and *Shigella* require T3SS for both host cell invasion as well as regulating the life cycles of the bacteria inside the host cell. *Salmonella* spp. have two distinct T3SSs: *Salmonella* pathogenicity island 1 (SPI-1) and *Salmonella* pathogenicity island 2 (SPI-2). Initially, the SPI-1 encoded T3SS is used for host cell invasion, during which the bacteria are engulfed into a vacuole inside the host cell, known as the *Salmonella* containing vacuole (SCV). The SPI-2 encoded T3SS is activated at a later stage during pathogenesis and its role is to regulate SCV trafficking, promoting the survival and replication of bacterial cells inside the host cell

Table 1.1: Bacteria secretion systems. Table source: ref. [2]

Secretion apparatus	Gram −ve or +ve	Substrate state	Number of membranes
Sec	Both	Unfolded	1
Tat	Both	Folded	1
T1SS	−ve	Unfolded	2
T2SS	−ve	Folded	1
T3SS	−ve	Unfolded	2-3
T4SS	−ve	Unfolded	2-3
T5SS	−ve	Unfolded	1
T6SS	−ve	Unknown	2-3
SecA2	+ve	Unfolded	1
Sortase	+ve	Folded	1
TS77	+ve	Folded	1-3

[23]. In extracellular pathogens such as EPEC, *Yersinia*, T3SS is used to modify the host cell cytoskeleton, which is a complex network of interlinking filaments and tubules that extend throughout the cytoplasm, as well as to manipulate the immune response of the host cell. In EPEC, the T3SS is used to attach to the intestinal epithelium of the host cell, and trigger cytoskeletal rearrangements to compromise the integrity of the epithelial barrier [24]. The T3SS is also used to avoid host cell immune activation by interrupting signalling in regulatory pathways such as nuclear factor-κB (NF-κB) as well as mitogen-activated protein kinase (MAPK) [25]. To better understand the T3SS let us look at the structure and function of various components that constitute this secretion pathway.

1.2 Structure and composition of the T3SS

The core of this macromolecular machine consists of a basal body: a stack of ring-shaped proteins which spans the inner and the outer membranes. The basal body, sometimes referred to as the base, allows for secretion of proteins that assemble other components of the T3SS as well as proteins that are involved in pathogenesis. At the end of the basal body, it has a narrow syringe-like structure or needle which extends from the outer membrane into the extracellular space. The basal body along with the needle is often referred to as the injectisome or the needle-complex. During pathogenesis, the injectisome serves as a transport channel and helps "inject" effector proteins into the host cells. The tip of the needle hosts translocator proteins, which interact with the host cell membrane and facilitate formation of a pore through which the effectors proteins are transferred [8].

The structure of the injectisome in various bacterial species has been studied in great detail using advanced techniques such as TEM. Genetic studies reveal that core proteins in the injectisome have significant similarities across all families as well as with those present in the flagellum [27, 28, 29]. The secretion and cellular translocation (Sct) prefix has been suggested as a unified nomenclature for the conserved components of the T3SS [30]. For the sake of convenience we will use this notation to describe the structure of the injectisome and flagellum in this section, and shall go back to using the traditional nomenclature in later sections and chapters. Table (1.2) enlists the core structural and functional components that are present in the flagellum and the T3SS across different bacterial species (IM: inner membrane and OM: outer membrane). The nine core proteins sharing sequence homology

Table 1.2: Core components of T3SS and their flagellar homologues. Table source: ref. [26]

Unified nomenclature	Function	*Yersinia* spp.	*Salmonella* spp. SPI-1	*Shigella* spp.	Flagellum
Basal body					
SctC	Secretin, OM ring	YscC	InvG	MxiD	–
–	Secretin pilotin	YscW	InvH	MxiM	–
SctD	IM ring	YscD	PrgH	MxiG	FliG
SctJ	**IM ring**	**YscJ**	**PrgK**	**MxiJ**	**FliF**
SctF	Needle	YscF	PrgI	MxiH	FlgE
SctI	Inner rod	YscI	PrgJ	MxiI	–
Export apparatus					
SctU	**Autoprotease**	**YscU**	**SpaS**	**Spa40**	**FlhB**
SctV	**Export gate**	**YscV**	**InvA**	**MxiA**	**FlhA**
SctR	**IM component**	**YscR**	**SpaP**	**Spa24**	**FliP**
SctS	**IM component**	**YscS**	**SpaQ**	**Spa9**	**FliQ**
SctT	**IM component**	**YscT**	**SpaR**	**Spa29**	**FliR**
SctQ	**C-ring**	**YscQ**	**SpaO**	**SsaQ**	**FliM/FliN**
ATPase complex					
SctN	**ATPase**	**YscN**	**InvC**	**Spa47**	**FliI**
SctL	**Stator**	**YscL**	**OrgB**	**MxiN**	**FliH**
SctO	Stalk	YscO	InvI	Spa13	FliJ
SctK	Cofactor	YscK	OrgA	MxiK	–
Translocators					
SctB	Pore	YopD	SipC	IpaC	–
SctE	Pore	YopB	SipB	IpaB	–
SctA	Tip	LcrV	SipD	IpaD	FliC
SctP	Ruler	YscP	InvJ	Spa32	FliK

across different species as well as the flagellar components are highlighted in boldfaced text. The basal body consists of series of rings of varying diameters stacked together to span the space between the inner and the outer membrane [31, 32, 33]. SctJ and SctD form concentric inner membrane rings, whereas SctC forms the outer membrane rings. Using cryo-EM, Worrall *et. al* measured the external diameter of the lower rings in *Salmonella typhimurium* (PrgH and PrgK) to be 24 nm [34]. Different rotational symmetries have been observed in the inner rings for different species, although the source of this variation is unclear at this point [35, 36]. The outer membrane ring SctC forms a double-layered β-barrel in the outer membrane [26]. The export apparatus is located below the basal body and consists of five membrane proteins; namely SctR, SctS, SctT, SctU, and SctV. SctU is an autoprotease, and there are conformational differences in the C-terminus of its pre and post-cleaved states. This conformational difference is relayed as a signal to stop the assembly of the needle by causing a switch in substrate specificity [37, 38, 39]. The cytoplasmic ring (C-ring), composed of SctQ, is present underneath the export apparatus; the C-ring and the ATPase together form the sorting platform for protein secretion [33]. The ATPase complex, which comprises the ATPase SctN, the stalk protein SctO, the stator protein SctL, and the cofactor SctK, is present beneath the C-ring. The stator SctL and stalk SctO help in anchoring the ATPase complex to the C-ring. The chief role of the ATPase complex is to energize the protein export [40, 41]. The width of the ATPase complex was first measured in the flagellum, and later in T3SS injectisomes in *Pseudomonas syringae*. The inner diameter of the ATPase complex is extremely narrow (about 2–3.8 nm) allowing secretion of only unfolded proteins [42, 43]. Figure (1.2) shows the key structural elements of the basal body and the C-ring and the export apparatus respectively.

Structural analysis of the basal body in *Salmonella typhimurium* shows the presence of an

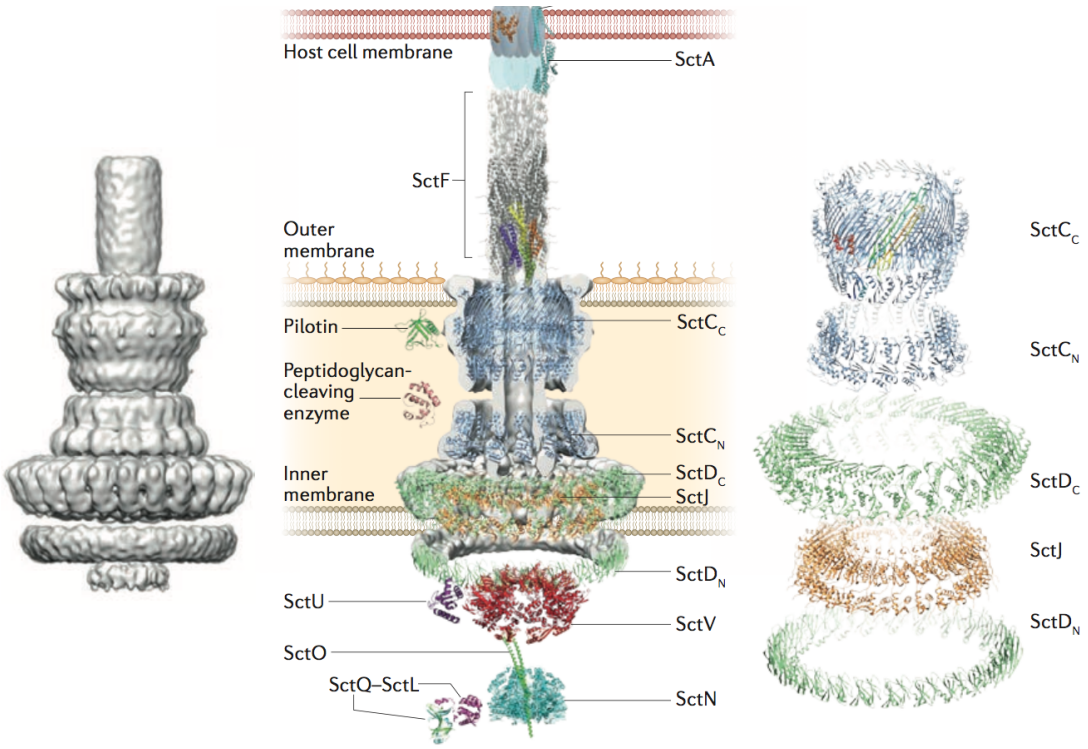


Figure 1.2: Structural elements of the basal body. Left: overall structure of the base, center: overlay of the protein sturcture of the components of the basal body, right: inner and outer membrane rings of the basal body. Figure source: ref. [26]

inner rod, composed of PrgJ (SctI), in the space between the inner and outer membranes. The exact dimensions of this structure are poorly characterized. However, mass spectrometry analysis of *Salmonella* SPI-1 T3SS shows that the stoichiometry of the inner rod protein is $\leq$6 [44]. The injectisome needle is a narrow cylindrical channel comprised of helical polymerization of hundreds of copies of a small protein, SctF, about 9 kDa [45, 46]. Loquet *et al.* observed that the helical parameters for the injectisome needle, composed of PrgI (SctF) in *Salmonella*, were 5.5 subunits per turn; which amounts to a 0.46 nm rise in the helix per subunit [47]. Structural studies reveal that the helical parameters of the needle protein MxiH (SctF) found in *Shigella flexneri* are identical to those of the needle protein PrgI (SctF) found in *Salmonella* [48]. In many animal pathogens there is a tip complex, composed of

SctA, which is present on the injectisome needles in the extracellular region. During the assembly process, secretion of the tip protein stops the secretion of needle subunits and prevents further growth of the needle. The tip complex is a pentameric structure composed of LcrV (SctA) in *Yersinia* spp. and of IpaD (SctA) in *Shigella* spp [49, 50]. Translocator proteins SctB and SctE, which are involved in pore formation in the host cell membrane to facilitate the transfer of effector proteins, are present at the end of the tip complex. Due to their small size, high resolution details of the tip complex and translocation pore are not yet available, however, there is experimental evidence that the tip complex functions as a scaffold for the assembly of the translocation pore in *Yersinia* and *Shigella* [49, 51, 52, 53]. While there is an overall similarity in the structures of T3SS injectisomes, there are some structural differences across species. For instance, in plant pathogens such as *Psuedomonas syringae*, the needle is comprised of a much longer structure, known as Hrp pilus [54]. In EPEC a long flexible filament, called the EspA filament, is present in place of the tip complex, and it is thought to act as a bridge across the mucus layer that separates the bacteria from the intestinal epithelial membrane [55, 56]. Figure (1.3) shows the structural differences in T3SS across different species.

The T3SS is also involved in assembly of the flagellum, a motile organelle that is found in many bacterial species. The basal body seen in the needle complex is also present in the flagellum, and many of the components are shared between the flagellar and injectisome basal bodies. In fact, recent genetic studies provide evidence that the non-flagellar T3SS evolved from the flagellum [57]. The inner membrane rings in the flagellum are referred to as membrane-supramembrane rings (MS-ring), whereas the outer membrane ring is known as the lipoprotein ring (L-ring). In addition to the inner and the outer membrane rings, there is a periplasmic ring (P-ring) present between the inner and the outer membranes in

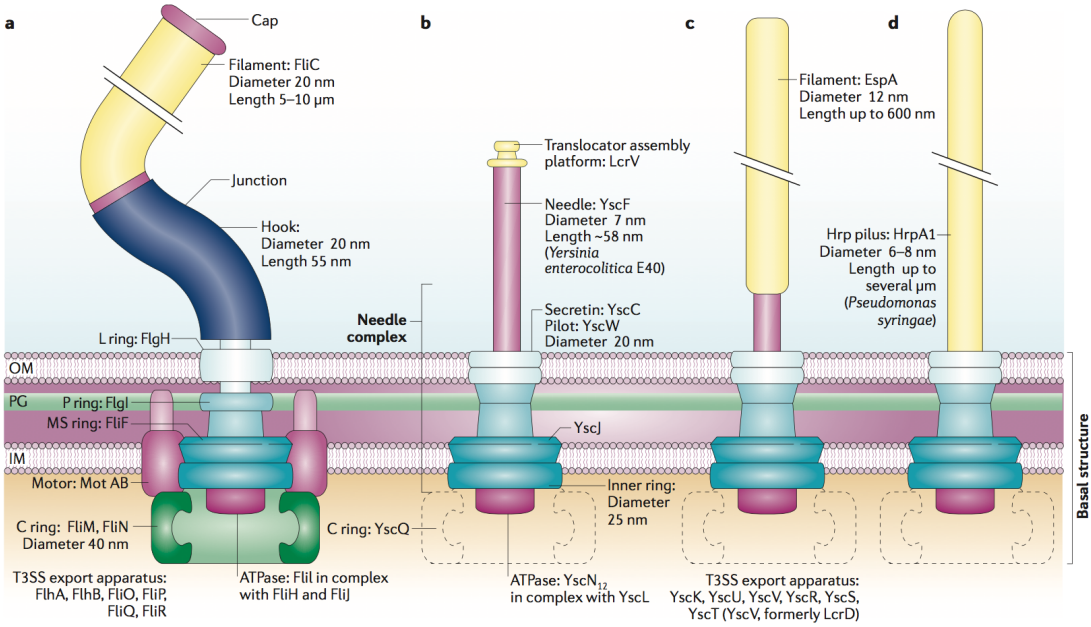


Figure 1.3: Structure of the flagellum and the injectisome. **a**: Schematic representation of the flagellum. **b**: Ysc injectisome. **c**: injectisome from enteropathogenic *E. coli*. **d**: injectisome from plant pathogens. Figure source: ref. [8]

the flagellar basal body [9]. The C-ring found in the injectisome basal body is also present in the flagellum and is much more prominent in the later [58]. In addition to the C-ring, the flagellum has a rotary motor, composed of MotA and MotB proteins, which is not present in the case of the injectisome. The rotary motor, which has an external diameter of about 45 nm, is energized by the flux of H^+ or Na^+ ions which are driven by the electrochemical gradient across the cytoplasmic membrane. The rotational speeds are as high as 100,000 revolutions per minute in some species. With the help of a structure known as the flagellar hook, this rotational motion is converted into a "whip" motion that propels the bacterium into forward motion [9, 59, 60]. The C-ring, together with the protein FliG, forms the switch complex which regulates the direction of the rotation [61]. The flagellar inner rod is composed of a variety of proteins: FlgF, FlgC, FlgB, FlgJ, FliE. However, not much is known

about the export properties of these proteins [62]. The flagellar hook is a curved cylindrical tube present on the outer membrane of the bacterial cell. As in the case of the injectisome needle, the flagellar hook consists of the helical polymerizations of hook subunits of protein FlgE (SctF) in *Salmonella enterica*. Even though hook subunits are much bigger ($\sim$ 45 kDa) than their needle counterparts, their helical parameters are identical to those of the needle subunits SctF(MxiH) in *Shigella flexneri* [63]. When the flagellar hook reaches the appropriate length, junction proteins are secreted, followed by the secretion of flagellin, which makes up the flagellar filament. The flagellar filament is an extremely long, thin, flexible structure, composed of as many as 20,000 flagellin subunits [9].

1.3 Length control mechanisms

From the structural details of the flagellum and the injectisome one can imagine that assembly of these nanomachines must be a highly regulated process. The assembly of the flagellum is extremely well understood and a detailed description of this process is provided in ref. [9]. The components of the basal body are secreted sequentially by means of the Sec pathway which was described earlier in this chapter. Two models have been proposed to explain the assembly of the T3SS core: the inside-out model, similar to the flagellar assembly, according to which the assembly of the export apparatus and the inner membrane rings takes place first, and the outside-in model, where the outer membrane rings are assembled first. There is evidence for inside-out assembly in *Salmonela* SPI-1 and EPEC, as well as for outside-in assembly in *Yersinia* spp [45, 64, 65]. The distal components, such as the hook and the filament, are assembled from protein secreted through the partially assembled T3SS channel. Given the structural resemblance of the flagellum and the injectisome, and the sequence ho-

mology in the proteins involved, the assembly mechanisms of these organelles are thought to be similar to one another[1]

To ensure smooth functioning and to optimize the efficiency of transport, it is important for the bacteria to regulate the lengths of these structures. If the injectisome is too short, then it fails to get past the protective layer present on the bacterial outer membrane known as the lipopolysaccharhide (LPS) layer. On the other hand, if the injectisome is too long then it results in more energy expended to transport proteins through longer channels. Furthermore, longer needles have a greater tendency to break from shear stress. In either case, a needle of inappropriate size is ineffective for pathogenesis. Similarly, a flagellar hook that is too long or too short does not possess the desired rotary properties to propel the bacteria. Experimental measurements for both, the flagellar hooks as well as the injectisome needles, show that the average lengths of these structures vary over a range of values across different species (see table (1.3)). This poses a natural question: how do the bacteria know when to stop growing these structures, especially when the structures themselves are outside the cell?

Various mechanisms have been proposed to explain the length regulation in the T3SS. Aizawa and co-workers observed that some C-ring mutants had shorter flagellar hooks. This observation led to the proposal of the cup model. The cup model suggests that the hook subunits accumulate in the cavity of the C-ring in the flagellar basal body. Once this cavity is full, the hook subunits are secreted to form the flagellar hook. After the completion of the hook, FliK protein is secreted and the interaction between FliK and the basal body causes the

[1]Note that in the case of the flagellum, cap proteins are secreted first, and in the absence of the cap proteins the hook subunits fail to polymerize and are lost in the extracellular environment. There is no needle-cap present in the T3SS injectisome, and the secretion of the tip protein takes place only upon completion of the injectisome needle [9].

Table 1.3: Average hook and needle lengths in different bacterial species. Table source: ref. [66]

Species	Hook/needle	Average length [nm]
Salmonella entrica serovar *Typhimurium*	Hook	55
Rodobacter sphaeroides	Hook	71
Yersinia enterocolitica	Needle	58
Yersinia pestis	Needle	41
Salmonella typhimurium	Needle	50
Shigella flexneri	Needle	45

bacteria to change the specificity of the substrate, from secreting hook subunits to secreting flagellin [67]. While the C-rings are present in the injectisome, they are not as prominent as in the flagellum [58, 68, 69, 70, 71]. Furthermore, there is very little experimental evidence in favor of the cup model. Apart from the cup model, there are two other length control mechanisms that have been proposed and studied in great detail over the past decade: the ruler mechanism and the substrate switching mechanism.

Ruler mechanism

The ruler mechanism has been proposed to explain length control in *Yersinia* needles. As per this model, a molecular ruler, YscP in *Yersinia*, is secreted through the needle. When the needle reaches the appropriate length, the C-terminus of the ruler molecule interacts with the gate protein, YscU, which cause the substrate specificity to switch, causing the base to stop

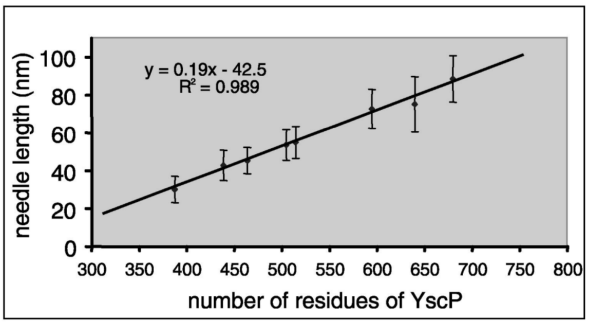
secreting needle subunits and secrete tip protein. In an experiment performed by Cornelis and co-workers, the length of the ruler was varied by deleting or inserting amino acid residues in the YscP molecule. They observed that there was a linear relationship between the ruler lengths and the average needle lengths, providing experimental evidence for the ruler mechanism (see Figure (1.4a)) [72]. The fact that the YscP molecules in *Y. enterocolitica* and *Y. pestis* differ by about 60 residues explains the difference in the average needle lengths in the two species [73]. Assuming an increase of about 0.19 nm per amino acid residue, Cornelis and co-workers estimated that the ruler domain of YscP would consist of 300–350 residues with over 150 residues for anchoring and signaling. It is known that the increase in the length can vary over a range of values and can be up to 0.4 nm per residue in extended polypeptide structures, but further evidence is required for the estimation of ruler length in *Yersinia*. The authors also observed that deletion of both the C-as well as the N-termini ($YscP_C$ and $YscP_N$) resulted in loss of length regulation, which lead them to hypothesize that ends of the ruler molecule act as anchors with $YscP_N$ attached to the tip of the needle and the $YscP_C$ attached to the basal body throughout the assembly process. In ref. [74] the authors provide further evidence for this anchoring by observing that even one molecule of YscP is sufficient to regulate the length in *Yersinia*.

The ruler mechanism was also proposed in the flagellar system. Using techniques such as gel filtration chromatography, multi-angle light scattering and analytical ultracentrifugation, Minamino *et al.* characterized the physiochemical properties of the ruler molecule FliK in *Salmonella*, and provided evidence that the binding interaction of $FliK_C$ to FlhB is structurally regulated by $FliK_N$, and occurs only after the hook has reached a certain predefined length [75]. Further evidence for the ruler mechanism in the flagellar system came from an experiment by Shibata *et al.* In this work the authors varied the length of the ruler molecule

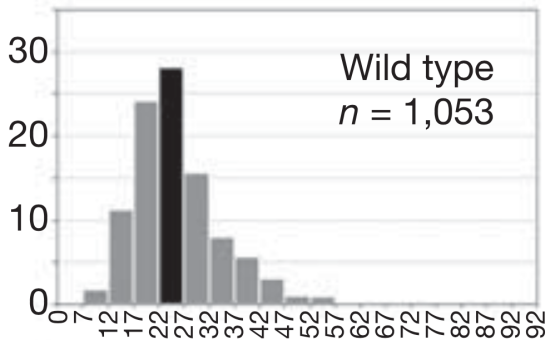
FliK and saw a linear increase in the hook lengths [66]. Recall that the width of the secretion channel is about 2–3.8 nm, making it highly unlikely for both the ruler molecule as well as the needle subunits to be present in the channel at the same time. This constraint becomes even more prominent in the flagellar system because of the fact that hook subunits are about 5 times bigger than the needle subunits. Keeping this in mind, a more dynamic version of the ruler model was proposed in length regulation of the flagellar hook. As per this model, the ruler protein FliK and hook proteins are secreted one at a time during the assembly process, the switch in the specificity depends on the speed at which FliK is being secreted. When the hook is too short, the tip of the hook, which binds to FliK_N, secretes the ruler molecule so rapidly that FliK_C does not have enough time to interact with the gate protein FlhB. It is only when the hook reaches the appropriate length that the secretion of FliK is sufficiently slow to allow for the the interaction between the C-terminus of FliK and the gate protein FlhB [76]. Once the ruler molecule has been secreted through the channel, it is lost in the extracellular space and cannot be reused for subsequent measurements. In chapter 3 we develop the mathematical model for the ruler mechanism, and we focus only on the dynamic version of the ruler mechanism.

Substrate switching mechanism

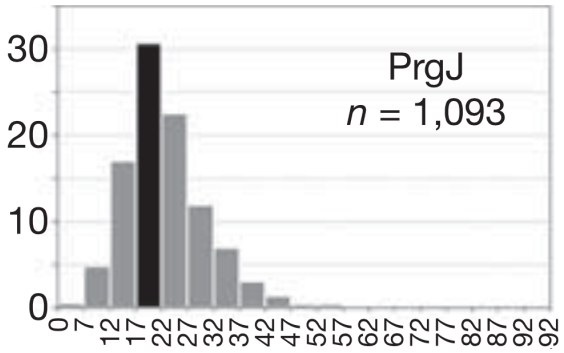
Based on the experiments performed with *Salmonella* and *Shigella*, a third model known as the substrate switching mechanism has been proposed to explain length regulation in the T3SS. In order to gain structural insights about the assembly of the T3SS, Unger and co-workers treated wildtype needles from *Salmonella* with a higher pH value. Western blot analysis revealed a significant reduction in the amount of the needle protein PrgI in the treated sample, but there was no difference in the amount of PrgJ between the treated



(a)



(b)



(c)

Figure 1.4: Experimental evidence for ruler and substrate switching mechanisms (a) shows the relationship between the average needle length and the ruler length in *Y. enterocolitica*. Figure source: ref. [72]. In (b) and (c) the needle length distribution of wildtype and PrgJ overexpressed mutants in *S. typhimurium* respectively. Figure source: ref. [77].

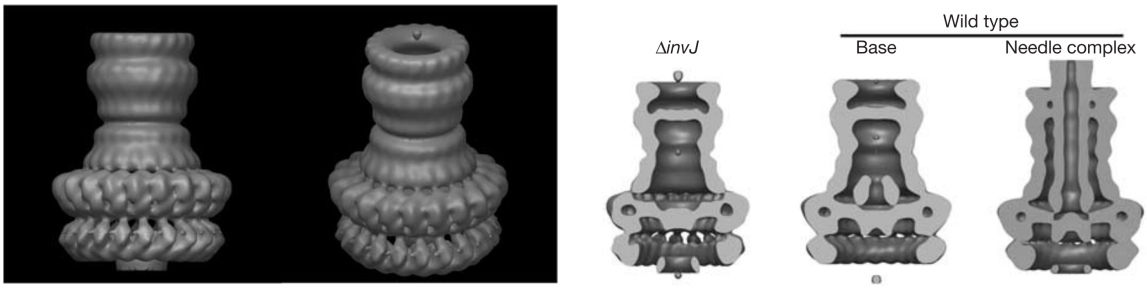
and untreated samples. The fact that PrgJ remained insensitive to changes in the external environment led to a conclusion that PrgJ must reside inside the basal body. Based on these findings, Galán and co-workers proposed the substrate switching model, according to which the assembly of the inner rod, composed of PrgJ, regulates the needle lengths [78]. According to this model, initially the inner rod and the needles are assembled at the same time, the completion of the inner rod causes the switch in substrate specificity and the base stops secreting needle proteins. Experimental studies showed that varying the amount of PrgJ had an impact on the needle lengths, thus providing evidence for the substrate switching model

(see figures (1.4b) and (1.4c), note the average needle lengths here are different from average lengths for *Salmonella* in table (1.3), and the sourceof this uncertainty is unclear). Galán and colleagues also observed that the lack of the protein InvJ resulted in abnormally long needles that were not firmly attached to the basal structure. Western blot analysis revealed that the inner rod protein PrgJ was absent in the needle complexes, and reconstruction of these $\Delta invJ$ mutants lacked the socket-like structure present in the wildtype basal bodies (see figure(1.5b)). Based on these observations, the authors concluded that the inner rod was not required for needle assembly, however the inner rod may be required for proper anchoring of the injectisome needle. The authors also characterized InvJ as a regulatory protein, that is not necessary for the assembly of the needle itself but is required to stabilize the conformation of socket in the basal body that allows the anchoring of the inner rod [77]. Experiments in *Shigella* suggest the presence of a protein, Spa47, which plays a similar role as InvJ in *Salmonella*, providing evidence for the assembly of *Shigella* needles as per the substrate switching mechanism [79, 80]. Note that as seen in table (1.2) InvJ protein is a functional homologue of protein YscP, and thus one might expect it to play the role of a molecular ruler in *Salmonella*, as does YscP in the case of *Yersinia*. In fact, recent experimental work by Wee and Hughes shows evidence for the ruler mechanism in the *Salmonella* SPI-1 injectisome. In this work the authors saw a linear increase in the needle lengths upon amino acid insertion in the N-terminus domain of InvJ in *Salmonella* needles. The authors provide an alternate explanation about the role of PrgJ protein–as a secretion competitor to needle protein PrgI, or as a protein that facilitates production of the ruler protein InvJ [73]. We can see that there is an ongoing controversy about the roles of PrgJ and InvJ proteins, and it is unclear if the current experimental methods would be able to resolve such issues.

1.4 Need for a quantitative framework

Over the past decade, competing models have been proposed to understand the length regulation in T3SS, and while detailed experiments have provided valuable insights on intricate aspects of these models, a full understanding about important features of these models has still not been achieved. Despite the rigorous experimentation on the structure and domain organization of the ruler protein, the actual length of the ruler domain still remains a mystery. Conflicting models have been proposed to explain length regulation in *Salmonella*. The role of some of its T3SS components, such as InvJ and PrgJ, is still controversial and more work needs to be done to reconcile the experimental observations. It is unclear whether experiments alone will suffice to fully characterize these models. Length measurements in a given bacterial species follow a distribution in lengths, and one might expect such length measurements to be normally distributed over a mean value. However, looking at the length distributions from various experiments in both the flagellar and the injectisome systems, one can see that the distributions often have a long tail, and the shape of these distributions has not been characterized mathematically. The effect of quantities such as the concentration of the needle subunits and number of ruler proteins on the average needle lengths is not understood quantitatively. Evolutionary pressures might have driven bacterial species to adopt different mechanisms for length control. However, the robustness of these mechanisms has not been explained. Keeping these issues in mind, we here formulate a mathematical framework for the substrate switching and the ruler mechanisms.

To date there has been no attempt to obtain a quantitative description of the substrate switching and the ruler mechanisms. In ref. [81] the author developed a simplified model for the ruler mechanism. However, the model fails to provide an explanation for the effect



(a) (b)

Figure 1.5: In the above figure (a) shows the surface rendering *S. typhimurium* basal body, (b) shows cropping of surface rendering of the $\Delta invJ$ mutant, the wildtype base and needle complexes in *S. typhimurium*. Figure source: ref. [77]

of crucial parameters, such as ruler length and copy numbers of ruler and needle proteins, on the needle lengths. To develop a mathematical model we considered the dynamics of the main constituent proteins involved in the assembly process; namely the base, needle proteins, the inner rod proteins, and ruler proteins. We considered the key processes involved in the assembly, such as synthesis, degradation, and maturation of the injectisome needle, and encoded the underlying dynamics into a system of ordinary differential equations (ODE's). Using the steady state solution of the ODE's, we also developed a statistical model that gives analytical expressions for needle length distribution and helps obtain a mathematical relationship between routinely measured quantities such as the average and the variance in lengths in experimentally relevant parameter regimes. To test the validity of the assumptions of our model, we performed numerical simulations based on the Doob-Gillespie algorithm. Not only do our models show excellent agreement with the available experimental data, but they also make testable predictions for both the substrate switching and ruler mechanisms.

Our mathematical model for the substrate switching mechanism predicts that there is a quadratic relationship between the variance and the average needle lengths, which is gov-

erned by only one free parameter, the number of inner rod proteins required to complete the assembly of the inner rod. Furthermore, comparing the results of our model with the available experimental data, our model predicted the number of inner rod proteins in *Salmonella typhimurium* to be around six, which was later found to be consistent with the experimental observations in ref. [44]. Our quantitative model for the ruler mechanism shows that an exact or error-free ruler is insufficient to explain the available experimental data and that a logistic ruler which has measurement uncertainty associated with it provides a more accurate description of the ruler mechanism. As per our model, there is no way of escaping the quadratic relationship between the average and the variance in the substrate switching mechanism, making it rather inefficient for regulating lengths in longer structures. An error-prone ruler, on the other hand, provides a more robust mechanism that offers a tight control over the length, even for extremely long structures. Our models provide insights on evolutionary trade-offs that might have forced bacteria to adopt different mechanisms for length regulation. Our work also suggests future experiments which that might help resolve the controversies arising from conflicting models [82, 83].

Chapter 2

Substrate Switching Mechanism

2.1 Introduction

In the previous chapter we discussed the two mechanisms for controlling the length of the needle that have been proposed, namely, the ruler and the substrate switching mechianisms. Cornelis and co-workers have posited that a dedicated protein might serve as a "molecular ruler." The idea in this case is that the C-terminus of the ruler protein can interact with the base, while the N-terminus can interact with the tip of the growing needle. As a result of this interaction, the length of the ruler is compared to the length of the needle, and when the ruler protein is stretched to its full length, this information is relayed to the base. Ruler proteins have been identified in various bacterial species and there is sufficient experimental evidence for their involvement in length regulation of the T3SS injectisome as well as the bacterial flagellum [66, 72, 73]. Experiments in *Salmonella* and *Shigella* show that there is an "inner rod" complex that is distinct from the outer needle complex; this inner rod spans the inner and outer membranes inside the base complex and is composed of a different protein from the needle itself. Completion of the inner rod leads to "substrate switching:" the

base stops exporting outer needle proteins and begins to secrete tip proteins and create a mature injectisome. Recent work from Hughes and co-workers also suggests the presence of a molecular ruler in *Salmonella*, and it is currently unclear which mechanism actually controls the length of *Salmonella* needles [73, 77, 78].

In this chapter, we develop a straightforward mathematical model for the substrate switching mechanism in *Salmonella*, which we validate using stochastic simulation. In the case where the majority of needles present on a cell at steady-state are functional (*i.e.* mature), analysis of the model revealed that there is a fixed relationship between the average and variance in needle length. Comparison of our results with the experimental data in Marlovits *et al.* indicates that a total of around six PrgJ proteins must bind to the inner rod in order to induce substrate switching [77]. This is consistent with recent experimental work by Zilkenat *et al.* on determination of stoichiometry of PrgJ proteins in *Salmonella* [44]. The distributions predicted by the model provide excellent agreement with experimental data with only a single free parameter. These findings provide quantitative support for the substrate switching model in *Salmonella*, and also suggest a set of straightforward experiments that would provide stringent tests of the mechanism in future work. Interestingly, the average lengths of needles in *Yersinia* are considerably longer than those observed in *Salmonella*, and the substrate switching model predicts that *Yersinia* needles should have a *much larger* variance in length than has been empirically observed. These results suggest that *Yersinia* and related species may have evolved a different mechanism (namely a ruler protein like YscP) to allow for more precise control over needle length when the needles themselves must be longer. Our work thus provides a mathematical framework for understanding the evolutionary pressures that have shaped length control mechanisms for this critically important virulence factor in bacteria.

2.2 Mathematical model

Ordinary differential equations

To construct the mathematical model for the substrate switching mechanism, we shall consider the dynamics of the three important constituent proteins; namely the immature base, the inner rod proteins and needle proteins; we shall represent these by symbols B, I and O, respectively. Each of these constituent proteins can be produced inside the cell, and each can be lost due to dilution from cell division. When a needle protein is exported through the base, the length of the needle increases by one unit. Similarly, the length of the inner rod increases upon export of an inner rod subunit. Figure (2.1) is a schematic diagram showing various stages during the assembly of the needle complex as per the substrate switching mechanism. The ordinary differential equations (ODE's) are constructed based on the following considerations:

(i) **Production:** The individual components in the system are synthesized in the cell at rates specified by parameters Q_B, Q_I and Q_O. This results in an overall increase in the number of proteins, *i.e.*

$$\frac{\mathrm{d}I}{\mathrm{d}t} = Q_I.$$

We would have a similar relation for the needle proteins and the bases as well. We assume that the Q's (and other rate parameters described below) are independent of time. As seen in the previous chapter, the molecular assembly of the base is a complex process in and of itself, however, in this model we represented the net effect as a constant production of bases.

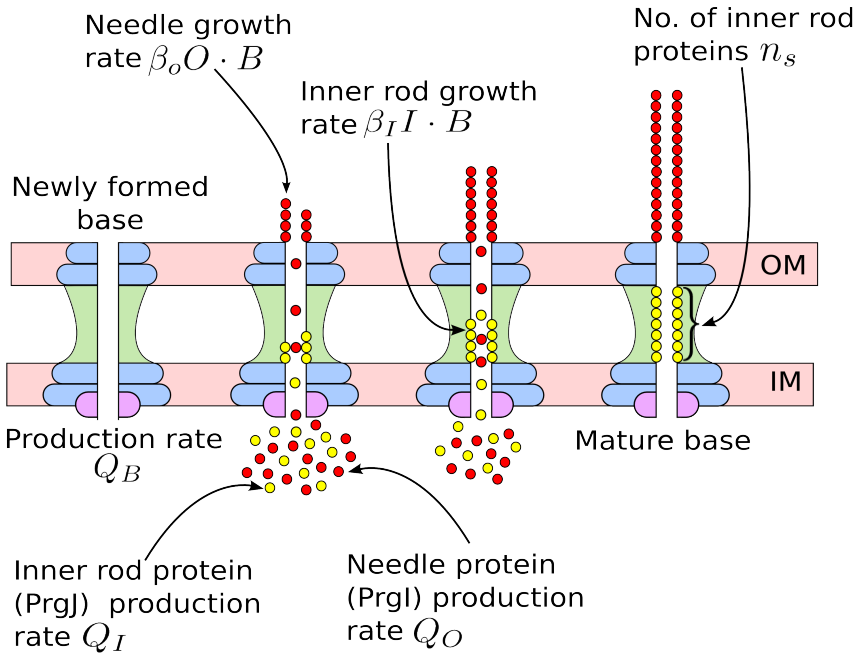


Figure 2.1: Schematic diagram showing various stages in the assembly of the needle complex as per the substrate switching mechanism.

(ii) **Degradation:** Dilution from cell division (and any additional active degradation processes) results in the effective loss of proteins. Following common convention as in refs. [84, 85], degradation/dilution is assumed to be first-order and depends on the rates λ_I, λ_O and λ_B. This results in an overall decrease in the number of proteins, *i.e.*

$$\frac{\mathrm{d}I}{\mathrm{d}t} = -\lambda_I I.$$

Note that the loss of a base would result in the loss not only of the base complex, but also any inner rod or outer needle proteins associated with it.

(iii) **Binding:** As mentioned in ref. [77], both the inner rod proteins and the outer needle proteins can bind to an immature base because substrate switching has not occurred yet. This binding is assumed to be a second-order reaction between the proteins and the base; the rate of this reaction is proportional to the product of the number of inner rod proteins (or outer needle proteins) and the number of bases and results in an overall

loss in number of proteins:

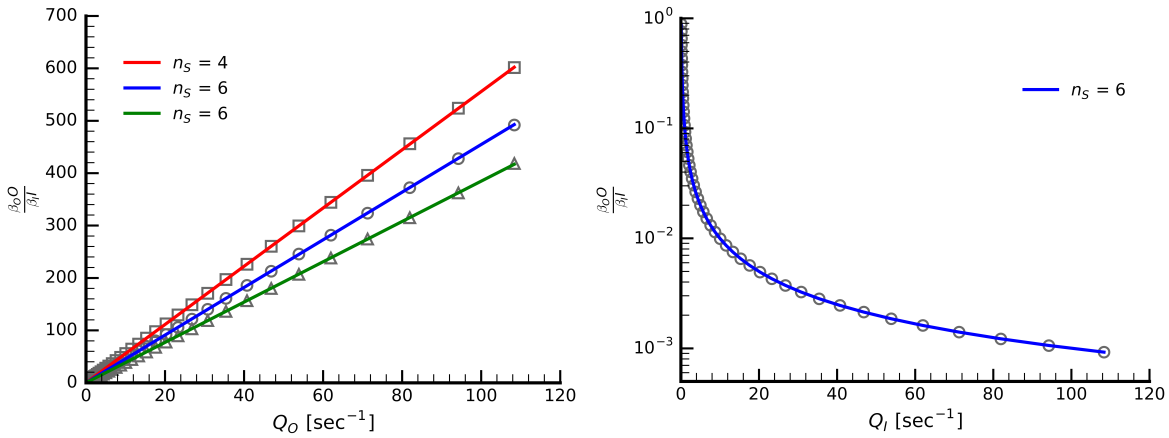
$$\frac{\mathrm{d}I}{\mathrm{d}t} = -\beta_I \cdot I \cdot B$$

The binding of an inner rod protein increases the length of the inner rod and the binding of a needle protein increases the length of the needle. Note that we do not explicitly consider the transport process for either inner rod or needle proteins. Also, we do not consider "unproductive" transport events, where, say, a needle protein might dissociate from the tip of the needle before the next needle protein binds. Since including this type of event would simply require a rescaling of the parameters in the model, we neglect these events without loss of generality. We also do not consider the possibility of a dissociated protein re-binding the needle, since the extracellular volume is likely so large that re-binding is highly unlikely.

(iv) **Substrate switching:** Substrate switching occurs when the number of inner rod proteins inside a base reaches a particular value n_S. Once substrate switching occurs, the base is no longer immature, which means that it cannot attach any more inner rod or outer needle proteins. For effective pathogenesis, it is crucial that the base is mature, since this is the only form that can secrete the tip and effector proteins. Note that the rate of association of inner rod proteins to an immature base is given by $\beta_I \cdot I \cdot B$ and for every n_S occurrences of association of inner rod proteins there is one substrate switching event, on average. In our simplified ODE model we thus approximate the maturation rate of bases by,

$$\frac{\mathrm{d}B}{\mathrm{d}t} = -\frac{1}{n_S}\beta_I \cdot I \cdot B.$$

Based on the above considerations we get that the following ODE's are representative of this system:



(a) (b)

Figure 2.2: Comparison between ODE solution and simulation

$$\frac{\mathrm{d}O}{\mathrm{d}t} = Q_O - \lambda_O O - \beta_O O \cdot B \tag{2.1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = Q_I - \lambda_I I - \beta_I I \cdot B \tag{2.2}$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = Q_B - \lambda_B B - \frac{1}{n_S}\beta_I I \cdot B. \tag{2.3}$$

The steady-state values of the number of the inner rod proteins, outer needle proteins and the immature bases can be found from equations (2.1), (2.2) and (2.3) as:

$$\bar{I} = \frac{1}{2\beta_I \lambda_I}\left(\sqrt{C^2 + 4n_S\beta_I\lambda_I\lambda_B Q_I} - C\right) \tag{2.4}$$

$$\bar{O} = \frac{Q_O\left(n_S\lambda_B + \beta_I\bar{I}\right)}{n_S\beta_O Q_B + \lambda_O\left(n_S\lambda_B + \beta_I\bar{I}\right)} \tag{2.5}$$

$$\bar{B} = \frac{n_S Q_B}{n_S\lambda_B + \beta_I\bar{I}}. \tag{2.6}$$

where $C = \beta_I\left(n_S Q_B - Q_I\right) + n_S\lambda_I\lambda_B$. Details of the stochastic simulations are discussed later in this chapter, figure (2.2) shows the comparison between results from theory and simulations.

Statistical model

To translate the average, "mean field" values for the number of inner rod, outer needle proteins and immature bases from equations (2.4)–(2.6) into the corresponding needle length distribution, we constructed a statistical model of the growth of the inner rod and needle. In this model, it is assumed that the system is already at the steady-state and each of the processes involved (production, binding and degradation) may be treated as uncorrelated sequential random events. For notational convenience we shall omit the bars (*e.g.* $\bar{B}$) and write I, O and B to indicate these steady-state averages below.

Imagine a base is synthesized at time $t = 0$. The probability that this particular base finishes its assembly such that the needle will achieve a length L (*i.e.* number of outer needle proteins bound) once it stops growing is:

$$P(L) = \int_0^\infty P(L;t)P(\text{stop};t)\mathrm{d}t, \tag{2.7}$$

where $P(L;t)$ is a conditional probability that the needle contains "L" proteins at time t, and $P(\text{stop};t)$ is the probability density that the needle stops growing at time t. Note that the rate at which the needle proteins bind to any given immature base is $\beta_O O$, and for a needle of length L, the number of consecutive protein bindings required to achieve this length is also L. Keeping this in mind and realizing that the process of an outer needle protein binding to an immature base is a Poisson process, we have:

$$P(L;t) = \frac{(\beta_O O t)^L e^{-\beta_O O t}}{L!}. \tag{2.8}$$

Note that we take I and O in this case to represent the mean field values from equations (2.4) and (2.5), respectively.

To calculate $P(\text{stop};t)$, we must consider the two independent mechanisms by which growth

of a particular outer needle might cease. Any given base might be degraded *before* its inner rod completes; alternatively, the inner rod might fully assemble (*i.e.* reach n_S proteins) before the base is degraded. The probability density of stopping at time t is simply the sum of these probabilities:

$$P(\text{stop};t) = P(\text{undeg};t)P(\text{complete};t) + P(\text{deg};t)P(\text{incomplete};t), \tag{2.9}$$

where $P(\text{undeg};t)$ and $P(\text{incomplete};t)$ denote the cumulative probabilities of the base remaining undegraded and the inner rod remaining incomplete until time t, respectively, and $P(\text{deg};t)$ and $P(\text{complete};t)$ denote the instantaneous probability densities that the base degraded and the inner rod is completed at time t, respectively. Base degradation occurs at rate λ_B, so the probability of the base remaining undegraded until time t is,

$$P(\text{undeg};t) = e^{-\lambda_B t}. \tag{2.10}$$

Similarly, the probability density that the base is degraded precisely at time t is,

$$P(\text{deg};t) = \lambda_B e^{-\lambda_B t}. \tag{2.11}$$

Once again, the process of each inner rod protein binding to an immature base is a Poisson process with rate $\beta_I I$. The density of the *waiting times* until the n_S^{th} Poisson event is thus given by an Erlang distribution:

$$P(\text{complete};t) = \frac{e^{-\beta_I I t}(\beta_I I)^{n_S} t^{n_S-1}}{(n_S-1)!}, \tag{2.12}$$

and the probability that the inner rod is incomplete is the cumulative probability that it remains incomplete until time t:

$$P(\text{incomplete};t) = e^{-\beta_I I t}\sum_{k=0}^{n_S-1}\frac{(\beta_I I t)^k}{k!}. \tag{2.13}$$

Using equations (2.8)–(2.13) in (2.7) we get:

$$P(L) = \int_0^\infty \frac{(\beta_O O t)^L e^{-\beta_O O t}}{L!}\left[\frac{e^{-\lambda_B t}e^{-\beta_I I t}(\beta_I I)^{n_S}t^{n_S-1}}{(n_S-1)!}\right.$$

$$+ \lambda_B e^{-\lambda_B t} e^{-\beta_I I t} \sum_{k=0}^{n_S-1} \frac{(\beta_I I t)^k}{k!} \Bigg] \mathrm{d}t \tag{2.14}$$

Let $\lambda = \beta_I I + \beta_O O + \lambda_B$, then

$$\begin{aligned} P(L) &= \frac{(\beta_O O)^L}{L!} \left[\frac{(\beta_I I)^{n_S}}{(n_S-1)!} \int_0^\infty t^{n_S+L-1} e^{-\lambda t} \mathrm{d}t + \lambda_B \sum_{k=0}^{n_S-1} \frac{(\beta_I I)^k}{k!} \int_0^\infty t^{k+L} e^{-\lambda t} \mathrm{d}t \right] \\ &= \frac{(\beta_O O)^L}{L!} \left[\frac{(\beta_I I)^{n_S}}{(n_S-1)!} \frac{\Gamma(n_S+L)}{\lambda^{n_S+L}} + \lambda_B \sum_{k=0}^{n_S-1} \frac{(\beta_I I)^k}{k!} \frac{\Gamma(k+L+1)}{\lambda^{k+L+1}} \right] \\ &= \frac{(\beta_O O)^L}{L!} \left[\frac{(\beta_I I)^{n_S}}{(n_S-1)!} \frac{(n_S+L-1)!}{\lambda^{n_S+L}} + \lambda_B \sum_{k=0}^{n_S-1} \frac{(\beta_I I)^k}{k!} \frac{(k+L)!}{\lambda^{k+L+1}} \right]. \end{aligned}$$

Using

$$y_O = \frac{\beta_O O}{\lambda}, \qquad y_I = \frac{\beta_I I}{\lambda}, \qquad \delta = \frac{\lambda_B}{\lambda}$$

in the above equation, we get;

$$P(L) = \frac{y_O^L}{L!} \left[\frac{(n_S+L-1)!}{(n_S-1)!} y_I^{n_S} + \delta \sum_{k=0}^{n_S-1} \frac{(k+L)!}{k!} y_I^k \right] \tag{2.15}$$

We can check for the normalization of $P(L)$:

$$\begin{aligned} \sum_{L=0}^{\infty} P(L) &= \sum_{L=0}^{\infty} \frac{y_O^L}{L!} \left[\frac{(n_S+L-1)!}{(n_S-1)!} y_I^{n_S} + \delta \sum_{k=0}^{n_S-1} \frac{(k+L)!}{k!} y_I^k \right] \\ &= \frac{y_I^{n_S}}{(n_S-1)!} \sum_{L=0}^{\infty} \frac{(n_S+L-1)!}{L!} y_O^L + \delta \sum_{k=0}^{n_S-1} \frac{y_I^k}{k!} \sum_{L=0}^{\infty} \frac{(k+L)!}{L!} y_O^L \\ &= \frac{y_I^{n_S}}{(1-y_O)^{n_S}} + \delta \sum_{k=0}^{n-1} \frac{y_I^k}{(1-y_O)^{k+1}} \\ &= z_I^n + (1-z_I) \sum_{k=0}^{n-1} z_I^k \;=\; 1 \end{aligned} \tag{2.16}$$

where

$$z_I = \frac{y_I}{1-y_O} = \frac{\beta_I I}{\beta_I I + \lambda_B}, \qquad z_O = \frac{y_O}{1-y_O} = \frac{\beta_O O}{\beta_I I + \lambda_B}, \qquad \text{and} \qquad \frac{\delta}{1-y_O} = 1 - z_I.$$

Note that we used

$$\frac{1}{m!}\sum_{L=0}^{\infty}\frac{(m+L)!}{L!}y^L = \frac{1}{(1-y)^{m+1}}$$

in the above calculation.

The average needle length is given by:

$$\begin{aligned}
\langle L\rangle &= \sum_{L=0}^{\infty} LP(L) \\
&= \sum_{L=1}^{\infty}\frac{y_O^L}{(L-1)!}\left[\frac{(n_S+L-1)!}{(n_S-1)!}\,y_I^{n_S} + \delta\sum_{k=0}^{n_S-1}\frac{(k+L)!}{k!}\,y_I^k\right] \\
&= \frac{y_I^{n_S}}{(n_S-1)!}\sum_{L=1}^{\infty}\frac{(n_S+L-1)!}{(L-1)!}y_O^L + \delta\sum_{k=0}^{n_S-1}\frac{y_I^k}{k!}\sum_{L=1}^{\infty}\frac{(k+L)!}{(L-1)!}y_O^L \\
&= \frac{y_I^{n_S}}{(n_S-1)!}\sum_{l=0}^{\infty}\frac{(n_S+l)!}{l!}y_O^{l+1} + \delta\sum_{k=0}^{n_S-1}\frac{y_I^k}{k!}\sum_{l=0}^{\infty}\frac{(k+l+1)!}{l!}y_O^{l+1} \\
&= y_O\left[\frac{n_S y_I^{n_S}}{(1-y_O)^{n_S+1}} + \delta\sum_{k=0}^{n_S-1}\frac{(k+1)y_I^k}{(1-y_O)^{k+2}}\right] \\
&= \frac{y_O}{1-y_O}\left[n_S z_I^{n_S} + \frac{\delta}{1-y_O}\sum_{k=1}^{n_S} k z_I^{k-1}\right] \\
&= z_O\left[n_S z_I^{n_S} + (1-z_I)\,\frac{\mathrm{d}}{\mathrm{d}z_I}\sum_{k=1}^{n_S} z_I^k\right] = z_O\frac{1-z_I^{n_S}}{1-z_I}.
\end{aligned} \tag{2.17}$$

Let

$$\epsilon = \frac{\lambda_B}{\beta_I I},$$

then

$$z_I = \frac{1}{1+\epsilon}\,, \qquad \frac{z_O}{1-z_I} = \frac{\beta_O O}{\beta_I I}\epsilon^{-1}$$

and

$$\langle L\rangle = \left(\frac{\beta_O O}{\beta_I I}\right)\frac{(1+\epsilon)^{n_S}-1}{\epsilon(1+\epsilon)^{n_S}}. \tag{2.18}$$

To calculate the standard deviation of L, we notice that

$$\langle L^2\rangle = \sum_{L=0}^{\infty} L^2 P(L) = \sum_{L=0}^{\infty} L(L-1+1)P(L) = \langle L\rangle + \sum_{L=0}^{\infty} L(L-1)P(L) \tag{2.19}$$

and

$$\sum_{L=0}^{\infty} L(L-1)P(L) = \frac{y_I^n}{(n-1)!} \sum_{L=2}^{\infty} \frac{(n+L-1)!}{(L-2)!}\, y_O^L + \delta \sum_{k=0}^{n-1} \frac{y_I^k}{k!} \sum_{L=2}^{\infty} \frac{(k+L)!}{(L-2)!}\, y_O^L. \tag{2.20}$$

Since

$$\sum_{L=2}^{\infty} \frac{(k+L)!}{(L-2)!}\, y_O^L = \sum_{l=0}^{\infty} \frac{(k+l+2)!}{l!}\, y_O^{l+2} = (k+2)! \frac{y_O^2}{(1-y_0)^{k+3}}, \tag{2.21}$$

$$\begin{aligned}
\sum_{L=0}^{\infty} L(L-1)P(L) &= \frac{y_I^{n_S}}{(n_S-1)!} \frac{(n_S+1)!\, y_O^2}{(1-y_0)^{n_S+2}} + \delta \sum_{k=0}^{n-1} \frac{y_I^k}{k!} \frac{(k+2)!\, y_O^2}{(1-y_0)^{k+3}} \\
&= \left(\frac{y_O}{1-y_O}\right)^2 \left[n_S(n_S+1) \frac{y_I^{n_S}}{(1-y_O)^{n_S}} \right. \\
&+ \left. \frac{\delta}{1-y_O} \sum_{k=0}^{n_S-1} (k+2)(k+1) \frac{y_I^k}{(1-y_O)^k} \right] \\
&= z_O^2 \left[n_S(n_S+1) z_I^{n_S} + (1-z_I) \sum_{k=0}^{n_S-1} (k+2)(k+1) z_I^k \right] \\
&= z_O^2 \left[n_S(n_S+1) z_I^{n_S} + (1-z_I) \frac{\mathrm{d}^2}{\mathrm{d}z_I^2} \sum_{k=0}^{n_S-1} z_I^{k+2} \right] \\
&= z_O^2 \left[n_S(n_S+1) z_I^{n_S} + (1-z_I) \frac{\mathrm{d}^2}{\mathrm{d}z_I^2} \left(\frac{z_I^2 - z_I^{n_S+2}}{1-z_I} \right) \right] \\
&= 2z_O^2 \frac{1 - (n_S+1) z_I^{n_S} + n_S z_I^{n_S+1}}{(1-z_I)^2}.
\end{aligned} \tag{2.22}$$

From equation (3.11),

$$z_I^{n_S} = \frac{z_O - (1-z_I)\langle L \rangle}{z_O}$$

and

$$\begin{aligned}
\sum_{L=0}^{\infty} L(L-1)P(L) &= 2z_O \left[\left(n_S + \frac{1}{1-z_I} \right) \langle L \rangle - n_S \frac{z_O}{1-z_I} \right] \\
&= \frac{2}{1+\epsilon} \left(\frac{\beta_O O}{\beta_I I} \right) \left[\left(n_S + 1 + \frac{1}{\epsilon} \right) \langle L \rangle - \frac{n_S}{\epsilon} \left(\frac{\beta_O O}{\beta_I I} \right) \right].
\end{aligned} \tag{2.23}$$

Therefore

$$\langle L^2 \rangle = \langle L \rangle + \sum_{L=0}^{\infty} L(L-1)P(L)$$

$$= \langle L \rangle + \frac{2}{1+\epsilon} \left(\frac{\beta_O O}{\beta_I I} \right) \left[\left(n_S + 1 + \frac{1}{\epsilon} \right) \langle L \rangle - \frac{n_S}{\epsilon} \left(\frac{\beta_O O}{\beta_I I} \right) \right] \tag{2.24}$$

and

$$\begin{aligned} \sigma^2 &= \langle L^2 \rangle - \langle L \rangle^2 \\ &= \langle L \rangle + \frac{2}{1+\epsilon} \left(\frac{\beta_O O}{\beta_I I} \right) \left[\left(n_S + 1 + \frac{1}{\epsilon} \right) \langle L \rangle - \frac{n_S}{\epsilon} \left(\frac{\beta_O O}{\beta_I I} \right) \right] - \langle L \rangle^2 \end{aligned} \tag{2.25}$$

2.3 Stochastic simulations

We treat the bases as individual, discrete "agents" in our simulations. There are two integers associated with each base; the first represents the number of inner rod proteins, and can take values from 0 to n_S, the second represents the number of needle proteins associated with that particular base, and can take any non-negative integer value. We maintain two separate populations of bases; the set of "immature" bases, with less than n_S inner rod proteins (call this set B), and "mature" bases, with exactly n_S inner rod proteins (call this set B'). The cardinality of these sets (*i.e.* the numbers of mature and immature bases) are represented as B and B', respectively, with the total number of bases $B_{\text{tot}} = B + B'$. We used two additional integers, I and O, to track the total number of inner rod and needle proteins.

The possible chemical reactions in this system, and their influences on the values of the variables described above, are summarized in figure (2.3). Note that we initialize the simulation with a set of completely immature bases that each have 0 inner rod and outer needle proteins. When a new base is synthesized during the simulation, we created a new immature base with 0 inner and outer proteins, and added this base to the immature pool. When a base

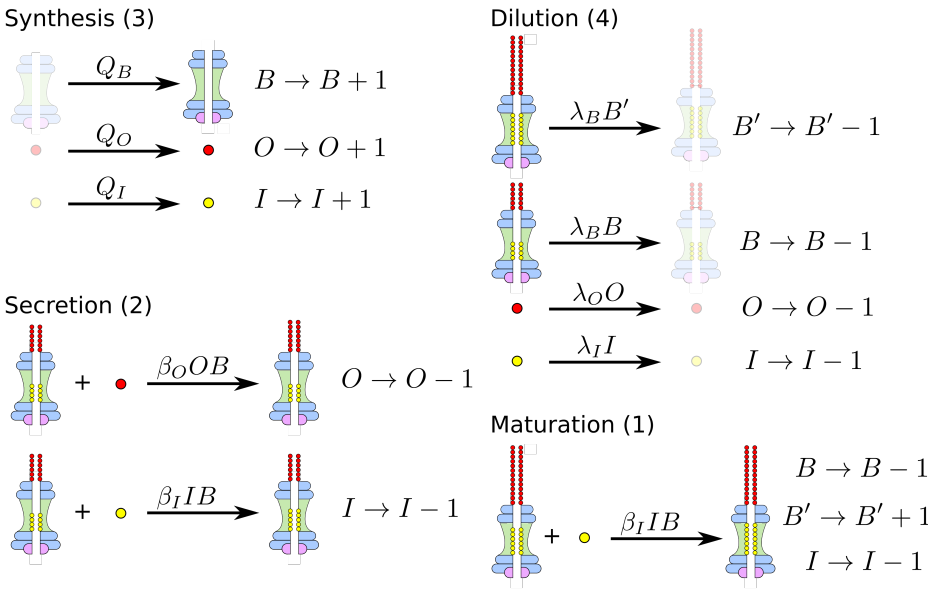


Figure 2.3: Individual processes involved in the assembly according to the substrate switching mechanism

is degraded, we chose one base at random from the set of total bases and removed it from the simulation; all bases, regardless of whether or not they are mature, have an equal probability of being chosen for degradation. Only immature bases can participate in binding reactions with inner rod or needle proteins, and when any given base binds to its "last" inner-rod protein (*i.e.* binds to an inner-rod protein and undergoes the transition from $n_S - 1$ to n_S inner rod proteins), that base becomes mature and is moved from the immature to mature pool.

A binding event in our model always results in an increase in the needle (or inner-rod) protein number associated with the base. As a result, we ignore a number of possible scenarios that might occur during needle assembly. For instance, a needle protein might dissociate from the needle; since the needle itself is a helix, the only protein that is likely to unbind is the one at the very tip. Alternatively, a needle protein might be exported by the base but simply never attach to the growing needle. Both of these scenarios represent "unsuccessful export," in that

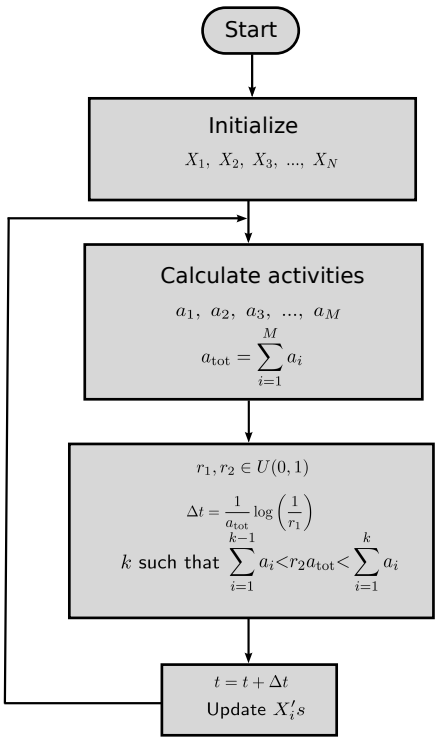


Figure 2.4: Flowchart of the Doob-Gillespie algorithm

the export event would reduce the number of needle proteins within the cell without a net increase in the needle length. This would mean that the $\beta_O O$ term in our deterministic equations (2.1)–(2.3) would no longer correspond to the $\beta_O O$ term in our statistical model, since the frequency of needle protein export would no longer be equal to the frequency of needle growth. We could thus define a separate rate constant for the statistical model, $\beta'_O < \beta_O$, to account for the fraction of export events that are unsuccessful. Since this would simply change the numerical relationship between the predictions of our deterministic model and the input to our statistical model, it would have no impact on the overall behavior of the system. Unsuccessful export could also lead to the accumulation of needle protein monomers in the extracellular space, which could bind to needles and extend them through a mechanism other than export. In our model, however, we assume that the extracellular volume is much larger

than the intracellular volume: as a result, re-binding events are likely to be very rare and are also neglected. In any case, re-binding would simply change the relative values of β'_O and β_O, so consideration of this effect would also not impact our results. As such, we neglect unsuccessful export events without a loss of generality.

We implement this simulation using the standard Doob-Gillespie approach for exact simulation of stochastic chemical kinetics [86]. Figure (2.4) shows the flowchart for the algorithm. The propensity or activity of any given reaction was calculated according to the functions over the arrows in figure (2.3). The rate constants in this case have exactly the same definitions as they do for the deterministic model described in the main text. The values for the parameters (*i.e.* Q's, β's, etc.) used in this study are summarized in table (2.1). All simulations were run until they achieved steady state; as a set of representative dynamics for various variables as they approach steady state.

2.4 Approximations for average and variance in lengths

Interestingly, the parameter ϵ in equation (2.18) plays a large part in determining the number of bases that are mature at steady state, since it compares the rate of base degradation/dilution (λ_B) to the frequency with which any given immature base will bind inner rod proteins ($\beta_I I$). When $\lambda_B \gg \beta_I I$ (equivalently $\epsilon \gg 1$), degradation dominates over inner rod protein binding, and most bases are degraded before they have enough time to bind n_S inner rod proteins. As a result, in this regime, the vast majority of bases are *immature* at steady state. Conversely, when $\lambda_B \ll \beta_I I$ ($\epsilon \ll 1$), then inner rod protein binding is much faster than degradation, and the majority of bases are *mature* at steady state. Injectisomes

are only functional when they are mature, and since each base complex represents a massive investment of energy in protein synthesis by the cell, we expect that the system has evolved towards a parameter regime where the majority of the bases are mature (and thus functional) at steady state. As such, we take $\epsilon \ll 1$ to be the relevant parameter regime for wildtype pathogenic bacteria.

Table 2.1: Parameter values used in simulation

Parameter	Value
Q_B	0.1 [molecules sec^{-1}]
Q_I, Q_O	0.1 –100 [molecules sec^{-1}]
λ_B, λ_I, λ_O	5×10^{-4} [sec^{-1}]
β_I, β_O	10^{-2} [$\text{molecules}^{-1}\text{sec}^{-1}$]
n_S	4–20

We find approximations for average length and variance of lengths in the biologically relevant case (where $\epsilon \ll 1$) and the case where dilution dominates over maturation ($\epsilon \gg 1$).

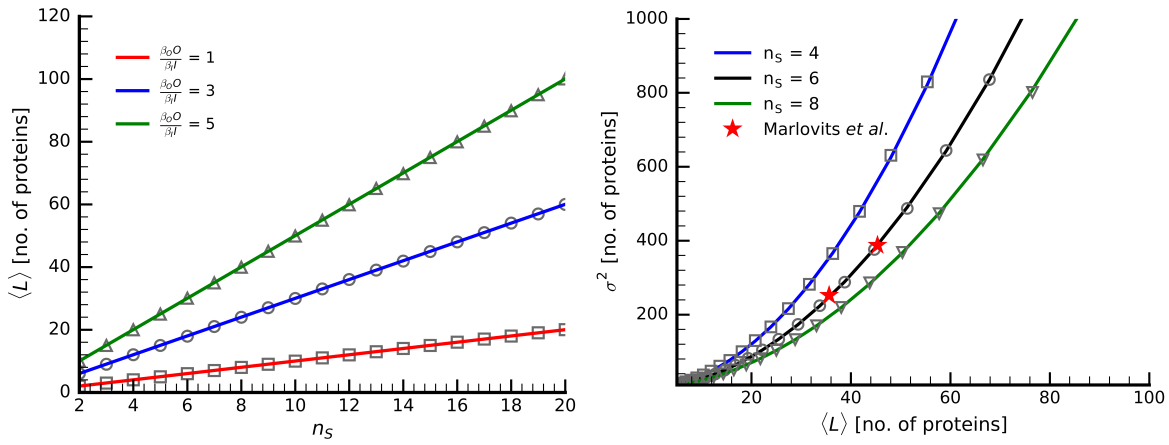
1. $\epsilon \ll 1$

From equation (2.18), we have

$$\langle L \rangle = \left(\frac{\beta_O O}{\beta_I I} \right) \frac{(1+\epsilon)^{n_S} - 1}{\epsilon (1+\epsilon)^{n_S}}.$$

Ignoring terms $\mathcal{O}(\epsilon^2)$ and higher in (2.18), the average needle length follows:

$$\langle L \rangle \simeq n_S \left(\frac{\beta_O O}{\beta_I I} \right). \tag{2.26}$$



(a) (b)

Figure 2.5: Average and variance in the $\epsilon \ll 1$ regime. Solid lines represent the predictions of the model, points represent the values from stochastic simulations. Data from ref. [77].

Similarly, keeping only zeroth order terms in ϵ in equation (2.25) we obtain the variance in the needle lengths as:

$$\begin{aligned}\sigma^2 &\simeq \langle L \rangle + n_S \frac{\beta_O O}{\beta_I I} \\ &= \langle L \rangle + \frac{\langle L \rangle^2}{n_S}.\end{aligned} \tag{2.27}$$

According to equation (2.27), the variance in the needle lengths should depend quadratically on the average needle length. Figures (2.5a) and (2.5b) show the comparison between the results from simulation and the predictions for average and variance as per equations (2.26) and (2.27). Furthermore, given a distribution of needle lengths we can predict the number of inner rod proteins required for substrate switching for any given bacterial species. Rearranging the terms in equations (2.27), we get:

$$n_S \simeq \frac{\langle L \rangle^2}{\sigma^2 - \langle L \rangle} \tag{2.28}$$

and using (2.28) in (2.26),

$$\frac{\beta_O O}{\beta_I I} \simeq \frac{\sigma^2}{\langle L \rangle} - 1. \tag{2.29}$$

Using equations (2.28) and (2.29) in the definitions of y_I, and y_O;

$$\begin{aligned} y_I &= \frac{\beta_I I}{\beta_I I + \beta_O O + \lambda_B} \\ &= \frac{1}{1 + \frac{\beta_O O}{\beta_I I} + \epsilon} \\ &\simeq \frac{\langle L \rangle}{\sigma^2 + \epsilon \langle L \rangle} \\ &\simeq \frac{\langle L \rangle}{\sigma^2} \end{aligned} \tag{2.30}$$

$$\begin{aligned} y_O &= \frac{\beta_O O}{\beta_I I + \beta_O O + \lambda_B} \\ &= \frac{\frac{\beta_O O}{\beta_I I}}{1 + \frac{\beta_O O}{\beta_I I} + \epsilon} \\ &\simeq \frac{\sigma^2 - \langle L \rangle}{\sigma^2 + \epsilon \langle L \rangle} \\ &\simeq 1 - \frac{\langle L \rangle}{\sigma^2}. \end{aligned} \tag{2.31}$$

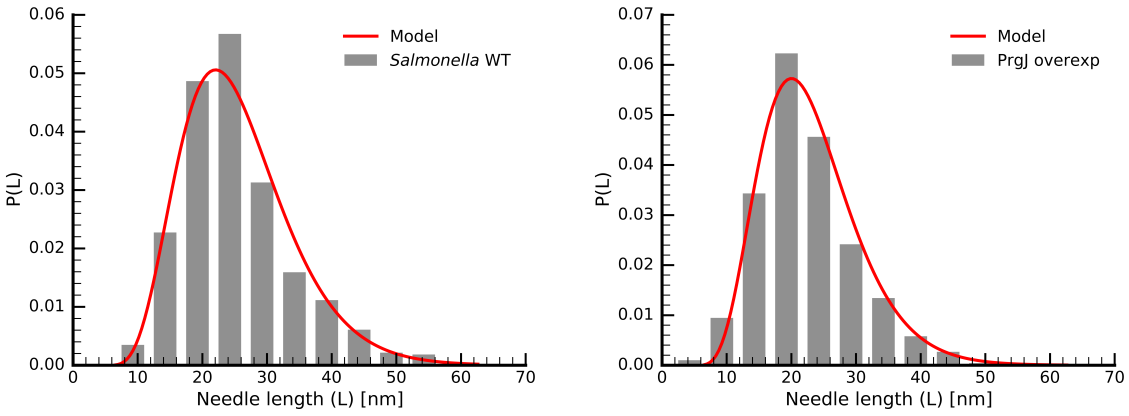
Thus for $\epsilon \ll 1$, $P(L)$ becomes,

$$\begin{aligned} P(L) &= \frac{y_O^L}{L!} \left[\frac{(n_S + L - 1)!}{(n_S - 1)!} y_I^{n_S} + \delta \sum_{k=0}^{n-1} \frac{(k + L)!}{k!} y_I^k \right] \\ &\simeq \frac{(n_S + L - 1)!}{L!(n_S - 1)!} y_I^{n_S} y_O^L \\ &\simeq \frac{(n_S + L - 1)!}{L!(n_S - 1)!} \left(\frac{\langle L \rangle}{\sigma^2} \right)^{n_s} \left(1 - \frac{\langle L \rangle}{\sigma^2} \right)^L \end{aligned} \tag{2.32}$$

The probability distribution of needle length in the $\epsilon \ll 1$ regime is a negative binomial distribution. Given a distribution of needle lengths from an experiment we can predict the number of inner rod proteins required for substrate switching for any given bacterial species using equation (2.32).

Comparison with experimental data

To further validate our findings, we compared them with experimental results obtained from ref. [77]. In this work, the authors were able to obtain two different distributions of needle



(a) (b)

Figure 2.6: Comparison of model with needle length distribution in ref. [77]

lengths in *Salmonella typhimurium*, one for "Wild Type" and the other for a mutant variant where the inner rod proteins (PrgJ) were overexpressed. We calculated the average needle length and the variance in the needle lengths from these distributions. Note that in the experiment, the needle lengths were measured in nanometers (nm), whereas in our work these lengths are represented in number of proteins, and hence we need a transformation that relates the two. As per the structural model of the needle in ref. [47], the height of the first subunit attached to the base is 6.04 nm and the increase in the height of the needle for a every outer needle (PrgI) binding is 0.46 nm, which gives us:

$$L \text{ [in nm]} = 0.46 \cdot L \text{ [in number of proteins]} + 6.04. \tag{2.33}$$

We then compared the predictions for the average and variance in needle length to those observed for the two experimental distributions (see figures (2.6a) and (2.6b)). Interestingly, both of these points lie on the same curve predicted by our model, indicating that the data for *Salmonella* is consistent with six inner rod proteins required for substrate switching (*i.e.* $n_S = 6$). Substituting $n_S = 6$ in equation (2.32), we obtained the analytical probability distribution of needle lengths and compared these with the experimental length distributions from ref. [77]. Note that we obtained excellent agreement between theory and experiment

with only one free parameter (n_S).

2. $\epsilon \gg 1$

For $\epsilon \gg 1$, the average length approximates to

$$\begin{aligned} \langle L \rangle &\simeq \frac{1}{\epsilon}\left(\frac{\beta_O O}{\beta_I I}\right) \\ &= \frac{\beta_O O}{\lambda_B} \end{aligned} \tag{2.34}$$

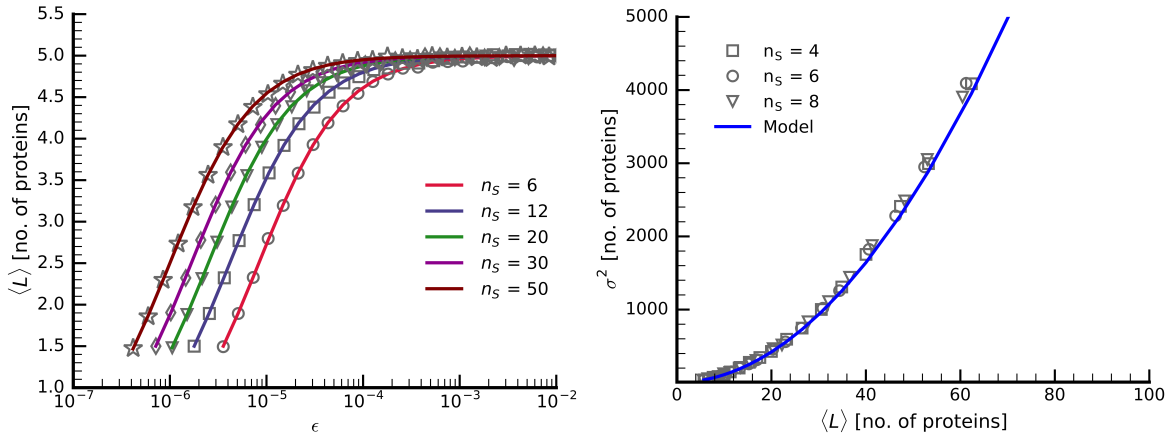
and the variance is

$$\begin{aligned} \sigma^2 &\simeq \langle L \rangle + \frac{1}{\epsilon^2}\left(\frac{\beta_O O}{\beta_I I}\right)^2 \\ &= \langle L \rangle + \langle L \rangle^2 \end{aligned} \tag{2.35}$$

Figure (2.7a) shows the average length against ϵ when β_I was varied in the range 10^{-2}–10^{-9}, with all other parameters as in table (2.1). We can see that the average length is independent of n_S and is determined by the rate parameters $\beta_O O$ and λ_B as predicted by equation (2.34). Figure (2.7b) shows a comparison between model and simulation variance against average length, and we can see that the variance is also independent of n_S for the $\epsilon \gg 1$ case. The ODE's, equations (2.1)–(2.3), do not always accurately represent the dynamics of the system. We use a different method, described below, to obtain the steady state values of I, O, and B used in calculating $P(L)$.

2.5 Exact ordinary differential equations

The term $\beta_I IB/n_S$ in equation (2.3) is an approximation and is not always valid, especially when a significant fraction of the bases are immature at steady state. When nearly 100% of



(a) (b)

Figure 2.7: Figure (a) shows average length versus ϵ when β_I is varied for different n_S values, (b) plots variance against the average length for $\epsilon \gg 1$. Solid lines represent the predictions of the model, points represent the values from stochastic simulations

the bases are mature at steady state, then for every n_S inner rod protein bindings, one base achieves maturation, on average. However, when most of the bases are immature, *i.e.* they have fewer than n_S inner rod proteins, at steady state, then the bases have to compete with one another for the inner rod (and needle) proteins. The system of ODE's that correctly represent this scenario is:

$$\frac{\mathrm{d}O}{\mathrm{d}t} = Q_O - \lambda_O O - \beta_O OB \tag{2.36}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = Q_I - \lambda_I I - \beta_I IB \tag{2.37}$$

$$\frac{\mathrm{d}B_0}{\mathrm{d}t} = Q_B - \lambda_B B_0 - \beta_I IB_0 \tag{2.38}$$

$$\frac{\mathrm{d}B_1}{\mathrm{d}t} = \beta_I IB_0 - \lambda_B B_1 - \beta_I IB_1 \tag{2.39}$$

$$\frac{\mathrm{d}B_2}{\mathrm{d}t} = \beta_I IB_1 - \lambda_B B_2 - \beta_I IB_2 \tag{2.40}$$

$$\vdots$$

$$\frac{\mathrm{d}B_{n_S-1}}{\mathrm{d}t} = \beta_I IB_{n_S-2} - \lambda_B B_{n_S-1} - \beta_I IB_{n_S-1}, \tag{2.41}$$

where B_i: is the number of bases with i inner-rod proteins attached to it and the total number of immature bases is:

$$B = \sum_{i=0}^{n_S-1} B_i. \tag{2.42}$$

The steady state solutions of equations (2.36) and (2.37) are;

$$O = \frac{Q_O}{\lambda_O + \beta_O B}, \tag{2.43}$$

and

$$I = \frac{Q_I}{\lambda_I + \beta_I B}. \tag{2.44}$$

To obtain B we need to solve equations (2.38)–(2.41).

$$B_0 = \frac{Q_B}{\lambda_B + \beta_I I}, \tag{2.45}$$

$$B_1 = \frac{\beta_I I B_0}{\lambda_B + \beta_I I} = \frac{Q_B \beta_I I}{(\lambda_B + \beta_I I)^2}, \tag{2.46}$$

$$B_2 = \frac{\beta_I I B_1}{\lambda_B + \beta_I I} = \frac{Q_B (\beta_I I)^2}{(\lambda_B + \beta_I I)^3}, \tag{2.47}$$

$$\vdots$$

$$B_{n_S-1} = \frac{\beta_I I B_{n_S-2}}{\lambda_B + \beta_I I} = \frac{Q_B (\beta_I I)^{n_S-1}}{(\lambda_B + \beta_I I)^{n_S}}. \tag{2.48}$$

Using equations (2.45)–(2.48) in equation (2.42)

$$B = \frac{Q_B}{\lambda_B + \beta_I I} \sum_{i=0}^{n_s-1} \left(\frac{\beta_I I}{\lambda_B + \beta_I I} \right)^i. \tag{2.49}$$

Substituting equation (2.49) into (2.44) gives a n_s^{th} order polynomial in I:

$$\lambda_I I + Q_B \sum_{i=0}^{n_s-1} \left(\frac{\beta_I I}{\lambda_B + \beta_I I} \right)^{i+1} - Q_I = 0. \tag{2.50}$$

We solved equation (2.50) numerically. Note that of the n_S possible solutions for the polynomial equation, only one solution for the value of I is real, and thus physically relevant. Once the steady state value of I is determined, the steady state values of the bases, B, and needle

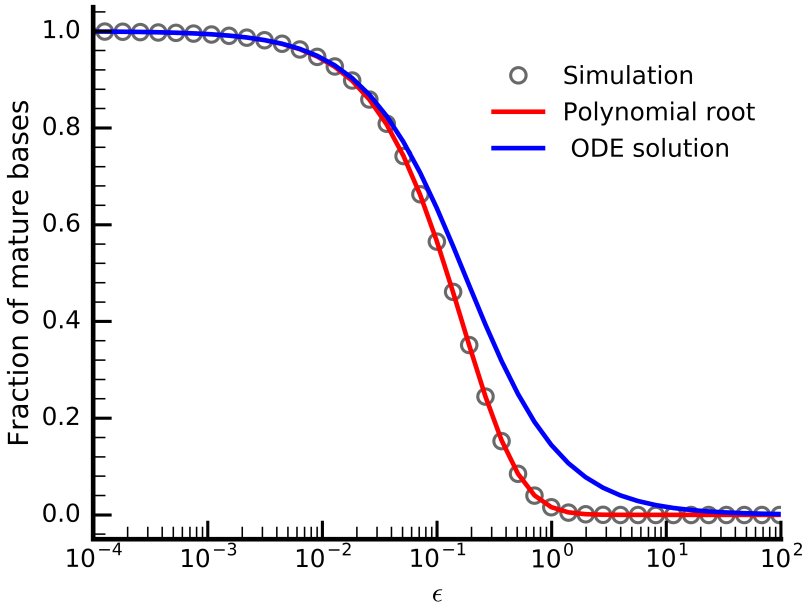


Figure 2.8: Comparison between the ODE solution, simulation, and solution of the polynomial equation (2.50) over a range of ϵ values.

proteins, O, can be easily determined. Figure (2.8) shows shows the discrepancy between the percentage of mature bases obtained from ODE solutions of (2.1)–(2.3) and the ones obtained from the polynomial solution of equation (2.50).

2.6 Discussion

Competing models have been proposed to explain how the length of the T3SS needle is controlled. While these models are generally considered to be in qualitative agreement with various experimental data sets [8, 72, 73, 77], there has been no attempt to use quantitative approaches to obtain and test predictions from these models in a rigorous way. In this chapter, we considered the substrate switching mechanism that has been proposed for length control in *Salmonella*, and constructed a mathematical model based on a straightforward

interpretation of the interactions thought to be important for this mechanism [8, 77]. The analytical results we obtained from this model are consistent not only with stochastic simulations, but also with available experimental data for *Salmonella typhimurium* [44, 77].

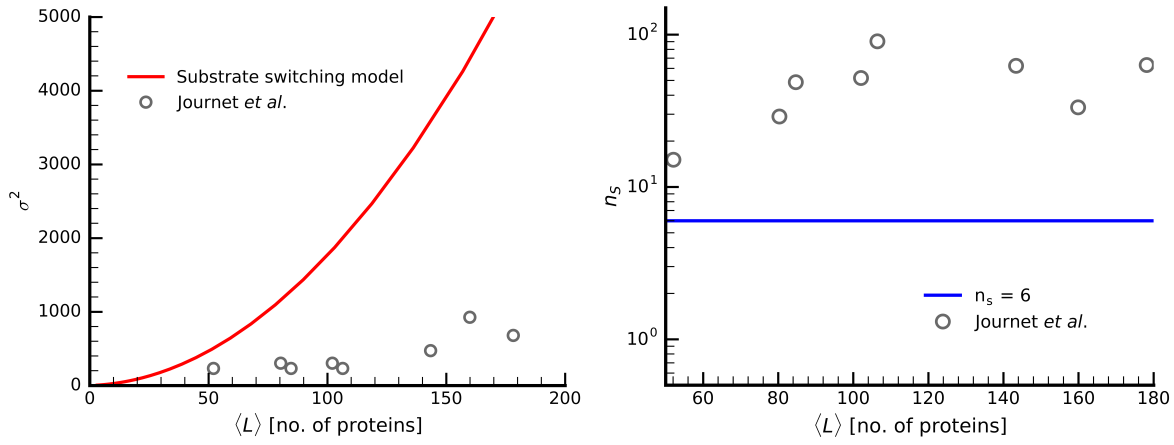
A major prediction of our model for $\epsilon \ll 1$ is that, for a given population of needles, the variance in the needle lengths should scale *quadratically* with the average needle lengths, parametrized by a consistent value of n_s (*i.e.* the number of inner rod proteins required for substrate switching, see equation (2.27)). This finding implies that the substrate switching mechanism may not be optimal for bacteria such as *Yersinia pestis*, which need to grow longer needles. If the length of the needles is regulated by substrate switching, larger values of the average length will result in very large variability, potentially impairing pathogenesis [72]. The quadratic relationship between the average and variance predicted by our model also suggests a natural set of experiments to provide further evidence for the substrate switching mechanism in *Salmonella* and other bacterial species. One could readily vary the synthesis rate of either PrgI or PrgJ (by using, say, "titratable" promoters whose activity can be controlled by the concentration of an exogenous ligand), measure the average and variance of the resulting needle lengths, and compare those results with the quadratic function predicted by our model (*e.g.* Figure (2.5b) with $n_S = 6$). Performing such an experiment would provide a very stringent test of the substrate switching mechanism.

Our simplified ODE model does not work for $\epsilon \gg 1$ and we need to solve a system of $n_S + 1$ ODE's (equations (2.36)–(2.41)) to obtain the steady state values. The assumptions of our statistical model are still valid and results of our model are in perfect agreement with stochastic simulations. For $\epsilon \gg 1$ our model predicts that the average length is independent of n_S and depends only on the needle protein export dilution rates. Figure (2.7a) shows

that the average lengths are distinct for different n_S values, and that the n_S dependence gradually recedes as we increase ϵ. For effective pathogenesis it is important that most injectisome needles achieve maturation, so the $\epsilon \gg 1$ parameter regime is not very exciting from an evolutionary perspective. However, $\epsilon \gg 1$ could be promising from a therapeutic perspective.

Our results also suggest that further structural studies could be helpful in determining the relative feasibility of the substrate switching model. EM structures of the base indicate that it is about 26.5 nm from top to bottom; the PrgJ portion of the inner rod takes up about 40%, or $\sim$ 10.5 nm, of this distance [78]. If we assume that PrgJ forms a helical structure identical to that of the outer needle, then according to equation (2.33) it would take $\sim$ 9 proteins to span that distance, which is not much larger than the $n_S = 6$ prediction made by our model. In order to fill that distance with six monomers, the PrgJ portion of the inner rod would have to adopt a somewhat more extended conformation than the needle complex, with $\sim$ 0.68 nm/monomer compared to 0.46 nm/monomer. Unfortunately, we currently do not have a higher-resolution structure of PrgJ nor is there a model of the inner rod structure akin to that of the needle [47]. Further structural studies could thus shed light onto whether the $n_S = 6$ prediction of the model is actually consistent with the number of PrgJ monomers found in the inner rod. Also, it is currently unclear how the information that the inner rod is complete is relayed to the other components of the base complex that ultimately decide which proteins to export and thus drive injectisome maturation. Atomic-resolution structures of more components of this complex will ultimately be necessary in order to understand the allosteric basis of the substrate switching mechanism.

We observed that the variances in these *Yersinia* needles are much smaller than the predictions for $n_s = 6$ according to equation (2.27) (see figure (2.9a)). Indeed, not only do the



(a) (b)

Figure 2.9: Experimental data for *Yersinia* do not agree with the results obtained from the substrate switching model.

variance and average length not satisfy the quadratic relationship with $n_s = 6$, there is no single n_S value with which the data are consistent. Applying equation (2.27) to these data, we observe a variety of n_S values scattered between ~ 10 to ~ 100 (see figure (2.9b)). As mentioned above, *Yersinia pestis* appears to have a molecular ruler (YscP) that regulates its needle lengths. Support for this model came from studies in *Yersina*, where it was shown that changes in the sequence length of YscP was strongly correlated with the average length of the resulting needles obtained from mutant cells [72]. Interestingly, Wee and Hughes recently provided evidence that changes in the length of the *Salmonella* homolog of YscP, InvJ, also causes concomitant increases in average length [73]. They argued that previous needle-length changes from PrgJ overexpression and deletion mutants in *Salmonella* were the result of changes in the frequency with which the length measurement is made by the ruler protein [73, 77]. In the next chapter, we shall discuss the details of the mathematical modeling for the ruler mechanism.

Chapter 3

Ruler Mechanism

3.1 Introduction

In the previous chapter, we constructed a mathematical model for the substrate switching mechanism. One of the key predictions of our model was that the variance in the needle lengths depends quadratically on the average length [82]. We found this relationship to be independent of all but one underlying parameter of the model, which is the number of inner proteins required to complete the inner rod. By comparing our results with the available experimental data, we predicted the number of inner protein subunits for completing the inner rod to be around six, which is consistent with recent experimental work on determination of stoichiometry of PrgJ proteins in *Salmonella* [44]. The quadratic relation between the variance and average length implies large variability in lengths, especially for longer needles. This makes substrate switching a potentially inefficient mechanism for growing longer structures. Furthermore, it is possible that different bacterial species have evolved distinct mechanisms for length regulation. The only way to test such hypotheses is by understanding the ruler mechanism in a similar quantitative framework, and to compare and contrast

the implications of the mathematical model with experimental data. While there is a rough agreement between needle length distributions and a simplified mathematical model of the ruler mechanism [81], to date there has been no quantitative exploration of how crucial parameters such as ruler length and the copy number of ruler proteins affect the average and variance in lengths.

In this chapter we shall develop a quantitative model for the ruler mechanism. Our initial model, in which the measurement of these structures takes place deterministically with the help of an exact ruler, predicts that variance should be *independent* of needle length and depend only on the number of ruler and needle proteins in the bacterial cell. Our analysis of experimental studies in which the authors obtained length distributions of these structures for a wide range of ruler lengths revealed a linear relationship between the variance and the average length [66, 72, 73]. A more realistic approximation of our model, where we introduce uncertainty in the measurement process by the ruler protein, results in a linear relationship between the variance in needles and the average needle lengths if the average length is increased by simply inserting additional amino acid residues into the ruler protein, which is what is done experimentally. Interestingly, if we allow the needle protein length to co-vary along with other parameters (such as the ruler protein concentration), it is possible to increase the needle length *without* increasing the variance, which is likely impossible in the substrate switching mechanism. Taken together, our findings suggest an interesting set of evolutionary trade-offs between these two mechanisms: the substrate switching mechanism allows for energy-efficient control of the lengths of smaller structures, while the ruler protein mechanism allows for robust control of the lengths of longer structures at a higher energy cost. Our results also suggest a very straightforward experimental program for further investigating the length control mechanism in a given bacterial species.

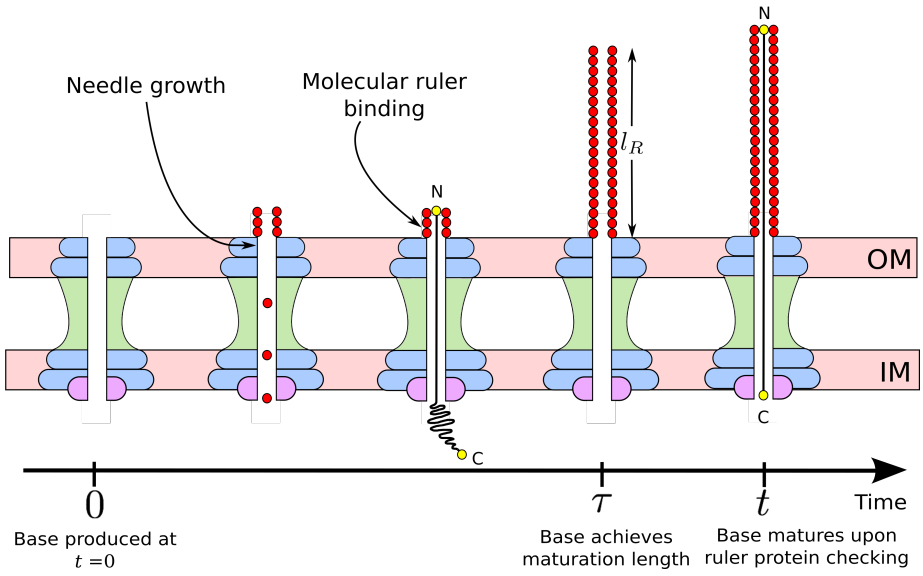


Figure 3.1: Schematic diagram showing various stages in the assembly of the needle complex as per the ruler mechanism.

3.2 Mathematical model

Ordinary differential equations

To construct the system of ODE's for the ruler mechanism we follow a procedure similar to what we did in the previous chapter for the substrate switching mechanism. Figure (3.1) depicts the mechanism of the needle growth in the T3SS injectisome using ruler proteins as per our initial mathematical model. Note that the figure only refers to the T3SS needle, but all the results described below are valid for the flagellar hook as well. Imagine a base being produced at time $t = 0$. During the assembly process, the ruler proteins and the needle proteins, represented by R and O respectively, are secreted the the base, represented by B, one at a time. The needle proteins polymerize in the extracellular region of the cell, resulting in growth of the needle itself. The needle length is represented by L, which is the distance between the surface of the outer membrane and the tip of the needle. At some time τ, after

a sufficient number of needle protein bindings, the needle achieves a length that is suitable for the switch in substrate specificity. In our model, we denote the effective length of the ruler protein by l_R, and when the needle reaches an appropriate length this facilitates the interaction between the ruler protein and the gate protein, switching the substrate specificity. Since the needle length has to be greater than or equal to the ruler length for maturation, all the ruler protein binding events that occur prior to the time τ are simply generating the necessary condition for maturation. As shown in figure (3.1), the next ruler protein binding event occurs at some (random) later time t, and it is at this time that the needle stops growing and the base becomes "mature." Since the maturation event occurs only when $L \geq l_R$, followed by a ruler protein checking event, for simplicity we ignored any uncertainties involved in the measurement process in our initial model, and as such we call it the *exact* ruler model. The binding rate of the ruler protein is β_R, which means the average time lapsed between two consecutive ruler protein bindings is $1/(\beta_R R)$. Thus, the average time required for maturation is:

$$\langle T \rangle = \tau + \frac{1}{\beta_R R}$$

If we assume the ruler protein binding and needle protein binding processes to be completely independent of one another then τ is :

$$\tau = \frac{l_R}{\beta_O O}$$

giving,

$$\langle T \rangle = \frac{l_R \beta_R R + \beta_O O}{\beta_O \beta_R O R}. \tag{3.1}$$

This means that, on average, after every $\langle T \rangle$ seconds a base achieves maturation and is lost from the pool of immature bases. Thus the ODE's for this system are:

$$\frac{\mathrm{d}B}{\mathrm{d}t} = Q_B - \lambda_B B - \frac{B}{\langle T \rangle}, \tag{3.2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = Q_R - \lambda_R R - \beta_R R B, \tag{3.3}$$

$$\frac{\mathrm{d}O}{\mathrm{d}t} = Q_O - \lambda_O O - \beta_O OB. \tag{3.4}$$

The solution to these ODE's gives the steady state values of O, R and B, which will be later used to calculate $\langle L \rangle$ and σ^2.

Statistical model

To develop a statistical model for the ruler mechanism we considered the dynamics of the main constituent proteins of the system, namely the base, the needle (or the hook) proteins and the ruler protein. The assembly of these structures involves the following key processes: synthesis or production of the constituent proteins in the bacterial cell, export of the needle subunits and the ruler proteins by the base, and dilution of these proteins from cell division.

We assume that the individual stochastic processes involved in the formation of the needle (namely the production, dilution and binding of the constituent proteins) are statistically independent, and we ignore any time delays that may arise due to synthesis or export of the proteins during the assembly process. We also assume that the system is already at steady state, *i.e.* the average values of B, O and R proteins are constant in time. Keeping these assumptions in mind, imagine a base being produced at $t = 0$. The probability that this base stops growing after achieving a certain needle length L can be written as:

$$P(L) = \int_0^\infty P(L \cap \text{stop}; t)\mathrm{d}t, \tag{3.5}$$

where $P(L \cap \text{stop}; t)$ is the probability that the base is of length L and it stops growing the needle at time t. This can happen if the base remains intact until time t and achieves maturation exactly at time t or that the base stays immature until and t and is lost due to

cell division at time t. Mathematically, this could be written as,

$$P(L \cap \text{stop}; t) = P(\text{undeg}; t)\ P(L \cap \text{mature}; t) + P(\text{deg}; t)\ P(L \cap \text{immature}; t), \tag{3.6}$$

$P(\text{undeg}; t)$ and $P(\text{deg}; t)$ are the same expressions as we had in case of substrate switching,

$$P(\text{undeg}; t) = \mathrm{e}^{-\lambda_B t}$$

$$P(\text{deg}; t) = \lambda_B\ \mathrm{e}^{-\lambda_B t},$$

where λ_B is the degradation rate of the base due to cell dilution.

$P(L \cap \text{mature}; t)$:

The process where a base has a needle of length L and is mature at time t can only occur if the base achieves the maturation length at an intermediate time τ and there is exactly one ruler protein binding event between τ and t;

$$P(L \cap \text{mature}; t) = \int_0^t \left(\mathrm{e}^{-\beta_O O \tau} \frac{(\beta_O O)^{l_R} \tau^{l_R - 1}}{(l_R - 1)!} \right) \beta_R R \mathrm{e}^{-\beta_R R (t-\tau)} \left(\mathrm{e}^{-\beta_O O (t-\tau)} \frac{\left[\beta_O O (t-\tau)\right]^{L - l_R}}{(L - l_R)!} \right) \mathrm{d}\tau \tag{3.7}$$

where β_O, β_R are the binding rates of the outer and ruler proteins, respectively; and O, R are the steady state values of the concentration of the outer and ruler proteins respectively. The third term in the equation (3.7) incorporates the fact that the base could have $(L - l_R)$ binding events after achieving the maturation length at τ and before the next ruler protein binding, at which point the base becomes mature.

$P(L \cap \text{immature}; t)$:

A base can remain immature if i.) it has fewer than l_R outer proteins bound to it at time t or ii.) it has achieved the maturation length at some intermediate time τ and there are no ruler protein bindings between the times τ and t.

$$P(L \cap \text{immature}; t) = \mathrm{e}^{-\beta_O O t} \frac{(\beta_O O t)^L}{L!} + \int_0^t \left(\mathrm{e}^{-\beta_O O \tau} \frac{(\beta_O O)^{l_R} \tau^{l_R - 1}}{(l_R - 1)!} \right)$$

$$\mathrm{e}^{-\beta_R R(t-\tau)}\left(\mathrm{e}^{-\beta_O O(t-\tau)}\frac{\left[\beta_O O(t-\tau)\right]^{L-l_R}}{(L-l_R)!}\right)\mathrm{d}\tau \tag{3.8}$$

Combining equations (3.5)–(3.8),

$$\begin{aligned}P(L) = \Theta(L-l_R)\int_0^\infty \mathrm{d}t\ \mathrm{e}^{-\lambda_B t}\int_0^t \mathrm{d}\tau\left(\mathrm{e}^{-\beta_O O\tau}\frac{(\beta_O O)^{l_R}\tau^{l_R-1}}{(l_R-1)!}\right)\beta_R R\mathrm{e}^{-\beta_R R(t-\tau)}\Bigg(\mathrm{e}^{-\beta_O O(t-\tau)}\\ \frac{\left[\beta_O O(t-\tau)\right]^{L-l_R}}{(L-l_R)!}\Bigg) + \Theta(l_R-L-1)\int_0^t \mathrm{d}t\ \lambda_B\ \mathrm{e}^{-\lambda_B t}\mathrm{e}^{-\beta_O O t}\frac{(\beta_O O t)^L}{L!}\\ + \Theta(L-l_R)\int_0^\infty \mathrm{d}t\ \lambda_B\mathrm{e}^{-\lambda_B t}\int_0^t \mathrm{d}\tau\ \left(\mathrm{e}^{-\beta_O O\tau}\frac{(\beta_O O)^{l_R}\tau^{l_R-1}}{(l_R-1)!}\right)\mathrm{e}^{-\beta_R R(t-\tau)}\\ \left(\mathrm{e}^{-\beta_O O(t-\tau)}\frac{\left[\beta_O O(t-\tau)\right]^{L-l_R}}{(L-l_R)!}\right).\end{aligned} \tag{3.9}$$

The Heaviside theta functions ensure that the term associated with mature needles does not have lengths $L < l_R$ and the term associated with immature needles does not have lengths $L \geq l_R$. We calculated the integrals in (3.9) using Mathematica (v11.0.0.0),

$$\begin{aligned}P(L) &= \Theta(L-l_R)\left[\frac{(\beta_O O)^L(\lambda_B+\beta_R R)}{(\lambda_B+\beta_O O)^{l_R}(\lambda_B+\beta_O O+\beta_R R)^{L+1-l_R}}\right]\\ &+ \Theta(l_R-L-1)\left[\frac{\lambda_B(\beta_O O)^L}{(\lambda_B+\beta_O O)^{L+1}}\right]\\ &= \frac{(\lambda_B+\beta_R R)(\lambda_B+\beta_O O+\beta_R R)^{l_R-1}}{(\lambda_B+\beta_O O)^{l_R}}\left(\frac{\beta_O O}{\lambda_B+\beta_O O+\beta_R R}\right)^L\Theta(L-l_R)\\ &+ \frac{\lambda_B}{\lambda_B+\beta_O O}\left(\frac{\beta_O O}{\lambda_B+\beta_O O}\right)^L\Theta(l_R-L-1)\\ &= \left(\frac{\lambda_B+\beta_R R}{\lambda_B+\beta_O O+\beta_R R}\right)\left(\frac{x_2}{x_1}\right)^{l_R}x_1^L\,\Theta(L-l_R)+\frac{\lambda_B}{\lambda_B+\beta_O O}\,x_2^L\,\Theta(l_R-L-1)\\ &= (1-x_1)\left(\frac{x_2}{x_1}\right)^{l_R}\Theta(L-l_R)\,x_1^L+(1-x_2)\,\Theta(l_R-L-1)\,x_2^L\end{aligned} \tag{3.10}$$

where

$$\begin{aligned}x_1 &= \frac{\beta_O O}{\lambda_B+\beta_O O+\beta_R R} = \frac{1}{1+(\lambda_B/\beta_O O)+(\beta_R R/\beta_O O)} = \frac{1}{1+z_O+z_O/z_R}\\ x_2 &= \frac{\beta_O O}{\lambda_B+\beta_O O} = \frac{1}{1+(\lambda_B/\beta_O O)} = \frac{1}{1+z_O}\end{aligned}$$

with

$$z_O = \frac{\lambda_B}{\beta_O O}\,, \qquad z_R = \frac{\lambda_B}{\beta_R R}\,, \qquad \frac{z_O}{z_R} = \frac{\beta_R R}{\beta_O O}$$

As we did in the case of substrate switching mechanism, we can check for normalization of $P(L)$:

$$\begin{aligned}\sum_{L=0}^{\infty} P(L) &= (1-x_1)\left(\frac{x_2}{x_1}\right)^{l_R} \sum_{L=l_R}^{\infty} x_1^L + (1-x_2) \sum_{L=0}^{l_R-1} x_2^L \\ &= (1-x_1)x_2^{l_R}\frac{1}{1-x_1} + (1-x_2)\frac{1-x_2^{l_R}}{1-x_2} \;=\; 1\end{aligned}$$

where

$$\sum_{L=l_R}^{\infty} x_1^L = \frac{x_1^{l_R}}{1-x_1} \qquad \text{and} \qquad \sum_{L=0}^{l_R-1} x_2^L = \frac{1-x_2^{l_R}}{1-x_2}$$

The average needle length is given by:

$$\begin{aligned}\langle L\rangle &= \sum_{L=0}^{\infty} P(L)L \;=\; (1-x_1)\left(\frac{x_2}{x_1}\right)^{l_R} \sum_{L=l_R}^{\infty} L x_1^L + (1-x_2) \sum_{L=0}^{l_R-1} L x_2^L \\ &= (1-x_1)\left(\frac{x_2}{x_1}\right)^{l_R} x_1 \frac{d}{dx_1}\left(\sum_{L=l_R}^{\infty} x_1^L\right) + (1-x_2)x_2 \frac{\mathrm{d}}{\mathrm{d}x_2}\left(\sum_{L=0}^{l_R-1} x_2^L\right) \\ &= (1-x_1)\left(\frac{x_2}{x_1}\right)^{l_R} x_1 \frac{d}{dx_1}\left(\frac{x_1^{l_R}}{1-x_1}\right) + (1-x_2)x_2 \frac{\mathrm{d}}{\mathrm{d}x_2}\left(\frac{1-x_2^{l_R}}{1-x_2}\right) \\ &= x_2^{l_R}\left(l_R + \frac{x_1}{1-x_1}\right) - l_R x_2^{l_R} + x_2\frac{1-x_2^{l_R}}{1-x_2} \\ &= \frac{x_2}{1-x_2} - x_2^{l_R}\left(\frac{x_2}{1-x_2} - \frac{x_1}{1-x_1}\right) \\ &= \frac{1}{z_O}\left(1 - \frac{x_2^{l_R}}{1+z_R}\right) \;=\; \frac{1}{z_O}\left[1 - \frac{1}{(1+z_O)^{l_R}(1+z_R)}\right] \qquad (3.11)\end{aligned}$$

where

$$\frac{x_1}{1-x_1} = \frac{z_R}{z_O(z_R+1)} \qquad \text{and} \qquad \frac{x_2}{1-x_2} = \frac{1}{z_O}$$

To obtain the variance in the needle lengths we calculate the second moment

$$\left\langle L^2 \right\rangle = \sum_{L=0}^{\infty} P(L)L^2 = \sum_{L=0}^{\infty} P(L)L(L-1) + \sum_{L=0}^{\infty} P(L)L = \sum_{L=0}^{\infty} P(L)L(L-1) + \langle L\rangle$$

where

$$\begin{aligned}
\sum_{L=0}^{\infty} P(L)L(L-1) &= (1-x_1)\left(\frac{x_2}{x_1}\right)^{l_R} \sum_{L=l_R}^{\infty} L(L-1)x_1^L + (1-x_2)\sum_{L=0}^{l_R-1} L(L-1)x_2^L \\
&= (1-x_1)\left(\frac{x_2}{x_1}\right)^{l_R} x_1^2 \frac{\mathrm{d}^2}{\mathrm{d}x_1^2}\left(\sum_{L=l_R}^{\infty} x_1^L\right) + (1-x_2)x_2^2 \frac{\mathrm{d}^2}{\mathrm{d}x_2^2}\left(\sum_{L=0}^{l_R-1} x_2^L\right) \\
&= (1-x_1)\left(\frac{x_2}{x_1}\right)^{l_R} x_1^2 \frac{\mathrm{d}^2}{\mathrm{d}x_1^2}\left(\frac{x_1^{l_R}}{1-x_1}\right) + (1-x_2)x_2^2 \frac{\mathrm{d}^2}{\mathrm{d}x_2^2}\left(\frac{1-x_2^{l_R}}{1-x_2}\right) \\
&= x_2^{l_R}\left[l_R(l_R-1) + \frac{2l_Rx_1}{1-x_1} + \frac{2x_1^2}{(1-x_1)^2}\right] + \frac{2x_2^2}{(1-x_2)^2} - x_2^{l_R}\left[l_R(l_R-1)\right. \\
&+ \left.\frac{2l_Rx_2}{1-x_2} + \frac{2x_2^2}{(1-x_2)^2}\right] \\
&= \frac{2x_2^2}{(1-x_2)^2} + 2x_2^{l_R}\left[l_R\left(\frac{x_1}{1-x_1} - \frac{x_2}{1-x_2}\right) + \frac{x_1^2}{(1-x_1)^2}\right. \\
&- \left.\frac{x_2^2}{(1-x_2)^2}\right] \\
&= \frac{2}{z_O^2} + 2x_2^{l_R}\left[-\frac{l_R}{z_O(1+z_R)} + \frac{z_R^2}{z_O^2(1+z_R)^2} - \frac{1}{z_O^2}\right] \\
&= \frac{2}{z_O^2}\left[1 - \frac{1}{(1+z_O)^{l_R}(1+z_R)}\left(2 + l_Rz_O - \frac{1}{1+z_R}\right)\right].
\end{aligned}$$

Since from $\langle L\rangle$ in equation (3.11), we have

$$\frac{1}{1+z_R} = (1+z_O)^{l_R}(1 - z_O\langle L\rangle),$$

$$\begin{aligned}
\sum_{L=0}^{\infty} P(L)L(L-1) &= \frac{2}{z_O^2}\left\{1 - \frac{(1+z_O)^{l_R}(1-z_O\langle L\rangle)}{(1+z_O)^{l_R}}\left[2 + l_Rz_O\right.\right. \\
&- \left.\left.(1+z_O)^{l_R}(1-z_O\langle L\rangle)\right]\right\} \\
&= \frac{2}{z_O^2}\left[1 - (2+l_Rz_O)(1-z_O\langle L\rangle) + (1+z_O)^{l_R}(1-z_O\langle L\rangle)^2\right].
\end{aligned}$$

Thus,

$$\begin{aligned}
\sigma^2 &= \left\langle L^2\right\rangle - \langle L\rangle^2 \\
&= \frac{2}{z_O^2}\left[1 - (2+l_Rz_O)(1-z_O\langle L\rangle) + (1+z_O)^{l_R}(1-z_O\langle L\rangle)^2\right] + \langle L\rangle - \langle L\rangle^2 \\
&= \frac{2}{z_O^2}\left[(1+z_O)^{l_R} - (1+l_Rz_O)\right]
\end{aligned}$$

$$
\begin{aligned}
& - \langle L \rangle \left\{ \frac{4}{z_O} \left[(1+z_O)^{l_R} - 1 \right] - (2l_R + 1) \right\} + \langle L \rangle^2 \left[2(1+z_O)^{l_R} - 1 \right] \\
= \;\; & a \langle L \rangle^2 - 2b \langle L \rangle + c \;=\; a \left(\langle L \rangle - \frac{b}{a} \right)^2 + \left(c - \frac{b^2}{a} \right)
\end{aligned} \tag{3.12}
$$

where a, b and c are:

$$
\begin{aligned}
a &= 2(1+z_O)^{l_R} - 1 \\
b &= \frac{2}{z_O} \left[(1+z_O)^{l_R} - (1 + \frac{l_R z_O}{2}) \right] - \frac{1}{2} \\
c &= \frac{2}{z_O^2} \left[(1+z_O)^{l_R} - (1 + l_R z_O) \right] \\
\text{and } \frac{b}{a} &= \frac{1}{z_O} - \frac{2 + (2l_R - 1) z_O}{2 z_O [2(1+z_O)^{l_R} - 1]}
\end{aligned}
$$

3.3 Approximations for average and variance in lengths

For the bacteria to have almost all needles mature, it should be able to produce and bind enough needle proteins, which means $\beta_O O \gg \lambda_B$ or $z_O \ll 1$. Furthermore, the needles should also be checked by attaching the ruler protein enough number of times or else it result in devastatingly large needles, which means $\beta_R R > \lambda_B$. At the same time it would be energetically inefficient for the bacteria to check the needle too many times, $\beta_R R \gg \beta_O O$. This leaves us with $z_R \simeq z_O$. So to obtain the expressions for average needle length and variance in the parameter space where most needles are mature we can ignore terms of $\mathcal{O}(z^2)$ or higher. From equation (3.11) we have

$$
\langle L \rangle = \frac{1}{z_O} \left[1 - \frac{1}{(1+z_O)^{l_R}(1+z_R)} \right]
$$

Ignoring terms of $\mathcal{O}(z^2)$ and higher

$$
\langle L \rangle \simeq \frac{1}{z_O} \left[\frac{(1 + l_R z_O)(1+z_R) - 1}{(1 + l_R z_O)(1+z_R)} \right]
$$

$$= \frac{1}{z_O}\left[\frac{l_R z_O + z_R}{1 + l_R z_O + z_R}\right].$$

Keeping the leading term in the denominator

$$\langle L \rangle \simeq l_R + \frac{z_R}{z_O}. \tag{3.13}$$

For the variance in length, we keep only the zeroth order terms in z for a, b, and c

$$\begin{aligned} a &= 2(1+z_O)^{l_R} - 1 && \overset{z_O \ll 1}{=} 1 \\ b &= \frac{2}{z_O}\left[(1+z_O)^{l_R} - (1 + \frac{l_R z_O}{2})\right] - \frac{1}{2} && \overset{z_O \ll 1}{=} l_R - \frac{1}{2} \\ c &= \frac{2}{z_O^2}\left[(1+z_O)^{l_R} - (1 + l_R z_O)\right] && \overset{z_O \ll 1}{=} l_R(l_R - 1) \end{aligned}$$

Using the approximate values of a, b, c and $\langle L \rangle$ in equation (3.12), we get

$$\sigma^2 \simeq \frac{z_R}{z_O} + \left(\frac{z_R}{z_O}\right)^2. \tag{3.14}$$

Comparison between model and stochastic simulation

In our derivation above, we made a number of simplifying assumptions (*e.g.* that the fluctuations in values of B, O and R are uncorrelated) that might not hold true in a bacterial cell. To test these assumptions, we performed stochastic simulations using the Doob-Gillespie algorithm [86]. In the simulations, the constituent proteins were treated as independent "agents" that interact with one another to form the needle (or the hook) [82, 84, 87, 88, 89]. Each base has an arbitrary number of needle proteins associated with it. We keep track of the number of unbound constituent proteins, number of mature and immature bases, as well as the number of outer proteins attached to a given base. The precise values of the parameters used in this model have not been determined experimently, so for the purposes of comparison with our analytical results we chose parameter values subjected to reasonable constraints. Note that in our model all lengths are measured in terms of number of proteins. We used

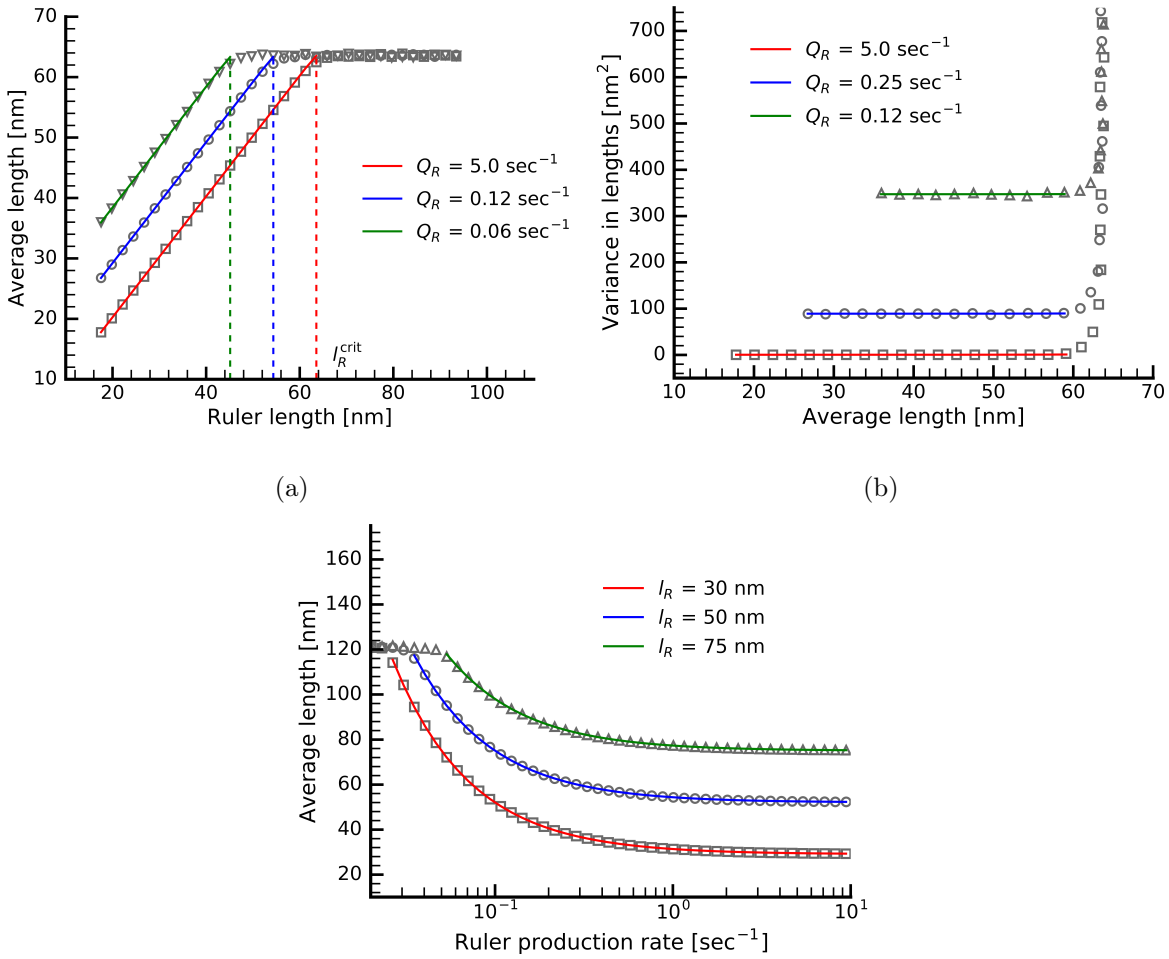


Figure 3.2: Comparison between the mathematical model and stochastic simulation for the average and variance as per the (exact) ruler mechanism.

the structural details in ref. [47] to convert the lengths to nanometers.

Figure (3.2) shows the comparison of the mathematical model and the stochastic simulation at steady state in terms of experimentally relevant quantities such as average length, variance in length and ruler production rate. Note that in figure (3.2) we do not plot the solid lines for the entire range of parameters because the assumptions of our model are not valid in these regions. In figure (3.2a), we observe a linear dependence of the average needle length on the ruler protein length for different ruler production rates, predicted as per equation

(3.13). We notice that the average needle length increases in response to increases in the ruler length only up to a certain critical value l_R^{crit}, beyond which it remains constant. The constant length is a consequence of the fact that the bases go from being mostly mature to mostly immature. For instance, in figure (3.2a) 100% of the bases are mature at $l_R = 10$ nm whereas only 40% of the bases are mature beyond l_R^{crit} for $Q_R = 5.0$ molecules/sec. The average length in this case is close to the ratio of production rates of the needle protein to that of the base (Q_O/Q_B) (in terms of number of proteins).

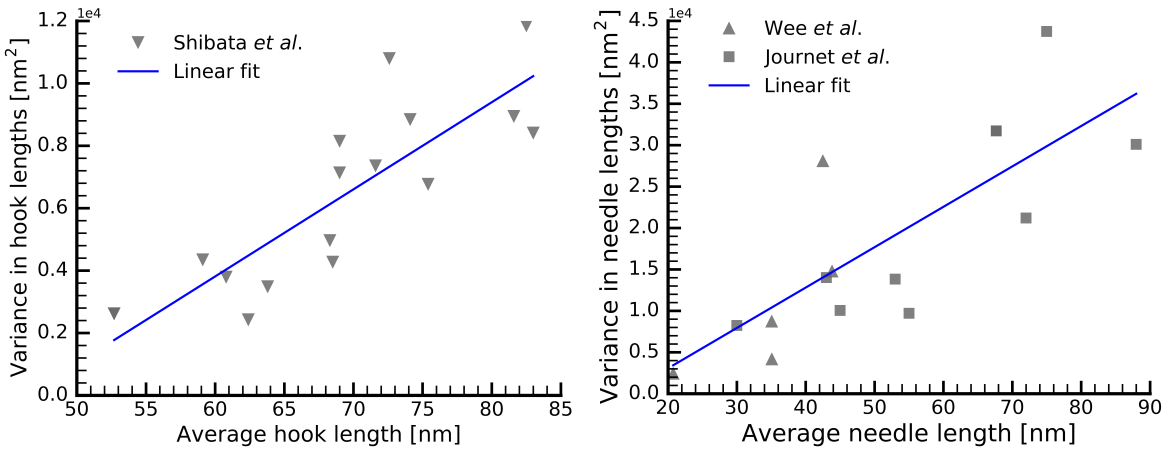
According to equation (3.14), the variance is independent of the average length and depends only on the number of ruler and needle proteins. In figure (3.2b), we can see that upon varying l_R, the variance remains constant with respect to the average length for a variety of ruler production rates. Figure (3.2b) also shows exceedingly large variances for extremely long ruler lengths. When the ruler proteins are extremely long, the percentage of immature bases increases, giving needles of varying lengths. The fact that the variance is independent of average length when most of the bases are mature makes the ruler model a unique mechanism. In the previous chapter we found that for the substrate switching mechanism the variance depends quadratically on the average length, and since the number of inner rod proteins required to complete the inner-rod is more or less constant for any given species, there is no way to increase the length without entailing the quadratic increase in variance. However, in the case of the (exact) ruler mechanism, the process of achieving the maturation length is completely deterministic. Thus the variance does not depend on the value of l_R, and depends only on the rate parameters, $\beta_O O$ and $\beta_R R$, making it possible to increase the average length by using longer rulers without changing the rate parameters of the ruler and needle proteins, leaving the variance unaltered (see equations (3.13) and (3.14)). Thus the ruler mechanism could provide a much more tightly regulated needle length distribution than

substrate switching.

Figure (3.2c) shows the change in average length when the ruler protein production rate is varied. For large values of ruler protein concentration, the checking rate increases and hence the needles achieve maturation at a length close to the ruler length. Decreasing the ruler protein production reduces the frequency of ruler checking, thus allowing the needles to grow beyond the ruler length before they become mature. At very low production rates of the ruler protein, there are so few rulers compared to the number of bases that most bases remain immature and the average needle length in the population no longer depends on l_R.

Experimental data

In ref. [72] the authors changed the ruler protein length by adding residues to the wild type ruler protein YscP in *Y. pestis* and saw a linear relation between the ruler length and the average needle lengths. A similar study was conducted by examining the flagellar hook lengths when the length of the FliK was varied by adding chimeric residues in ref. [66]. More recently, researchers have identified InvJ (in *S. enterica*), a protein homologous to YscP as well as FliK, and studied the effect of varying its length on the average length of T3SS needles [73]. The key result of these experiments was that the average length increased linearly with the length of the molecular ruler, giving strong evidence for the ruler mechanism. We obtained the variance of the needle and hook lengths from the data in these studies. Figures (3.3a) and (3.3b) show that the variance is not independent of the average length, but rather increases linearly with the average length. Our regression analysis confirms that there was no statistical significance for a quadratic relationship between these two quantities. Although a linear relationship between variance and needle length allows a tighter control over needle lengths as compared to a quadratic relationship, this experimental data is not consistent with



(a) (b)

Figure 3.3: Data for the flagellar hook in ref. [66] in figure (3.3a) and for T3SS injectisome in refs. [72, 73] in figure (3.3b) shows statistical significance for a linear relationship between the variance and the average length.

the exact ruler model.

3.4 Ruler length with an error-prone ruler

One of the assumptions in our initial model was that the interaction of the C-terminus of the ruler protein with the gate protein occurs at a fixed length. In reality, this interaction depends on conformational changes that arise as a consequence of the folding of the ruler protein, resulting in uncertainties in the measurement process. To incorporate this effect in our simulation, we allowed the length of the ruler involved in every binding interaction to fluctuate in such a way that the maturation probability obeys a logistic distribution [81]. This variability in the effective ruler length makes it possible for the base to achieve maturation for a needle length that might have been too short for maturation in the exact ruler model. This effect is not straightforward to incorporate in the analytical calculation for the probability distribution of needle lengths, and thus in the analysis that follows we focus on

simulation results.

Figure (3.4) shows the average and variance in lengths for a logistic ruler with a fixed error in length. In figure (3.4a) we observe that varying the length of the error-prone ruler (keeping all other parameters fixed) leads to a linear relationship between the average length and the ruler length, which is what was observed in the case of the exact ruler as well. In figure (3.4b) we can see the change in average lengths when the ruler production rate is varied. For lower production rates, only a few rulers are being produced in comparison with the number of bases available, giving us the same average lengths as obtained using the exact ruler. But when we increase the ruler production rates, the average lengths in the case of a logistic ruler decrease more rapidly than they did in the case of an exact ruler. This is because, at higher production rates, even though maturation at small values of L is less likely to occur, there are enough ruler checking events to ensure maturation even for needles with L significantly less than l_R. Note that this effect is prominent even at moderate values of Q_R. In figure (3.4c) we observe that varying the length of the error-prone ruler (same conditions as in figure (3.4a)) leads to a linear relationship between the variance and the average length. A logistic ruler allows for maturation events at much smaller needle lengths, giving a large number of needles whose lengths are less than l_R. This skews the distribution of needle lengths resulting in larger variances. Note that variance obtained through the simulation is smaller than that observed in the experiments, which could be due a number of experimental uncertainties that were not incorporated in our simulations.

As discussed earlier, for an exact ruler the variance in length is independent of the ruler length itself, and depends only on the needle and ruler protein concentrations and export rates. We should note that figures (3.4a) and (3.4c) represent a typical experimental situa-

tion in which the ruler protein length is changed *via* deletion of amino acid residues in the protein, or insertion of new residues into a specific region [66, 72, 73]. In this scenario, we find that the variance increases linearly with the average needle length, which is also what is observed experimentally (Figures (3.3) and (3.4c)). During the course of evolution, however, it would be possible for both the length of the ruler protein and the synthesis rates of the ruler and needle proteins to vary simultaneously. We thus tested whether it would be possible to find parameter sets where the increase in variance seen in figure (3.4c) could be offset by varying other parameters. Interestingly, we found that it is indeed possible to increase the average length of the needles *without* significantly changing the variance (figure (3.4d)). While more complex than the case with an exact ruler, it is thus nonetheless possible for an error-prone ruler to exhibit an independence of the average and variance in the length distribution.

3.5 Discussion

The T3SS and flagellar hook are large and complex nanomachines that are crucial to bacterial cell function, pathogenesis and adaptation to changing environments. One key aspect of these homologous structures is the large extracellular channel that must be constructed for both of them. The lengths of these channels have to be extremely precise; if they are too short or too long, that may negatively influence the corresponding function. Recall that if the T3SS needles are too short then they would be unable to get past the lipopolysaccharide (LPS) layer present in the extracelluar region of the bacterium, thus failing to inject the effector proteins into the host cell. On the other hand if the needles are too long, then they will likely break from shear stress or lead to energetically inefficent transport of proteins through an

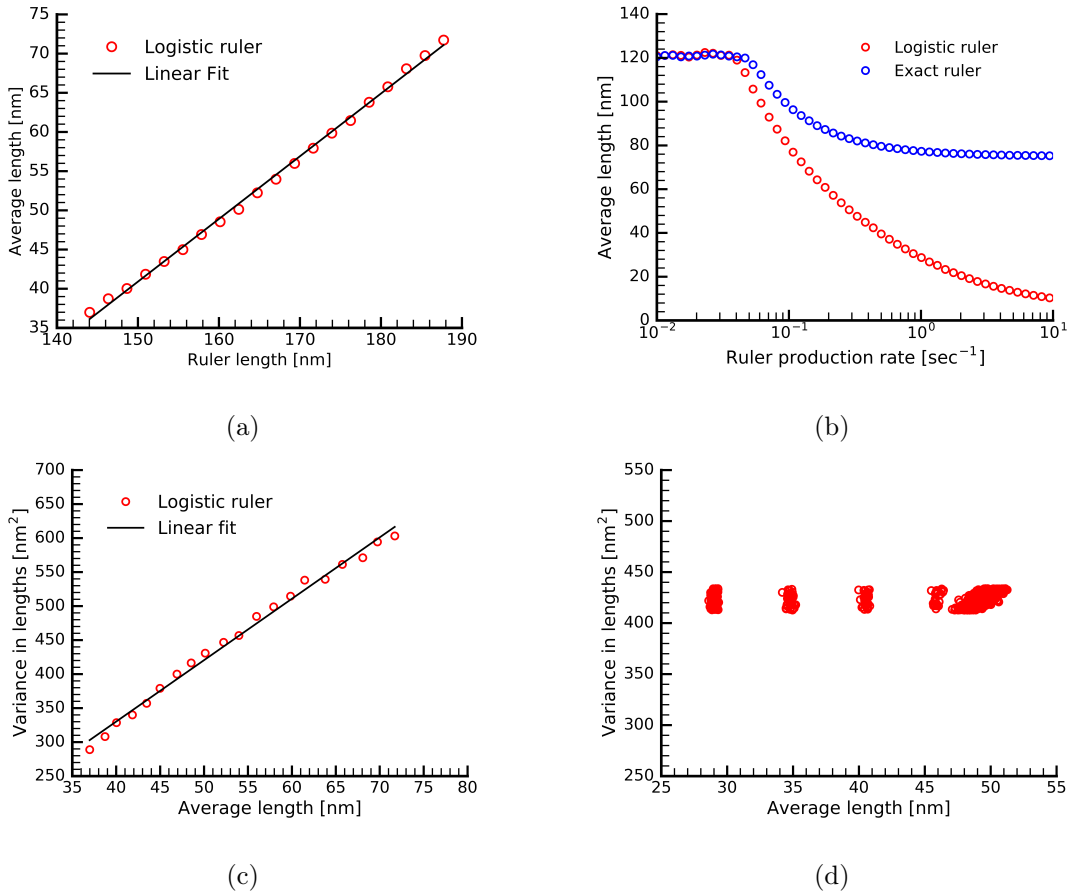


Figure 3.4: Figure (3.4a) shows that there is a linear relationship between average needle lengths and ruler length,. Figure (3.4b) shows that the average length asymptotically to zero with an increase in the ruler production. Figure (3.4c) shows that a logistic ruler is consistent with a linear relationship between the variance and the average needle lengths. Figure (3.4d) shows that it is possible to increase the average length without increasing the variance.

excessively long channel. Similarly, a flagellar hook of an inappropriate size might lead to poor mechanical properties for locomotion.

The substrate switching and ruler mechanisms are the two most popular and well-studied explanations for length control in these structures. The substrate switching mechanism requires completion of the inner rod present inside the base in order to form a mature needle, whereas according to the ruler mechanism, the bacteria "measure" the lengths of these struc-

tures with a dedicated ruler protein. While there is some experimental evidence for these mechanisms, until recently several important questions remained unanswered. For instance, is there a quantitative consistency of the proposed mechanism with experiments? What are the advantages, from an evolutionary perspective, to adapt different length control mechanisms? What future experiments could be done to further investigate these mechanisms?

In the previous chapter, we constructed a mathematical model for length control in the substrate switching mechanism and compared our predictions with the available experimental data for *Salmonella*. Our model was able to explain the distribution in the needle lengths obtained in ref. [77], and we were also able to predict the number of PrgJ proteins required to complete the inner rod and found that our prediction was consistent with data from mass spectrometry [44]. In the current chapter we developed a mathematical model for the ruler mechanism. Interestingly, we found that an exact ruler, which is perhaps most commonly envisioned for length measurement according to the ruler mechanism, does not explain the experimental data, and that a logistic ruler, which has a measurement uncertainty associated with it, is consistent with the available experimental data [66, 72, 73]. Interestingly, an error-prone ruler allows bacterial cells to regulate the variance to a desired value by adapting the ruler length and ruler and needle proteins synthesis rates simultaneously. This allows the ruler mechanism to exert a greater degree of control over the length of the structure than the substrate switching mechanism, particularly when the structures themselves are long. This higher degree of control, however, comes at a fairly high energy cost. In particular, once ruler proteins are secreted, they are lost, and obtaining a low variance requires many ruler protein “measurements" for every base that becomes mature. In contrast, the substrate switching model involves a constant number of inner-rod proteins for each base, and if this number is low (around six or so, as predicted by our model and observed in experimental data

[44, 82]), then the energy cost of control in this case is relatively low. We thus might expect the substrate switching mechanism to be favored as a low-energy control strategy for short structures, whereas the ruler protein mechanism might be favored for longer structures where the lower cost of the substrate switching mechanism would be outweighed by the massive increase in variance that it entails as the needles or hooks become longer.

Our quantitative models suggest several experiments that can understand such trade-offs in energy cost versus decreased variance in lengths. For instance, researchers have studied the effect of varying the length of ruler proteins such as YscP (in *Yersinia*) and FliK (in *Salmonella*) on the length of needles and flagellar hooks respectively, so an experiment where one changes the concentration of the ruler proteins by means of titratable promotors for the wild type as well as for ruler length variants of these ruler proteins would help find evidence for the ruler mechanism (See figure (3.2c) and (3.4c)). A similar experiment where one changes the concentration of the inner-rod protein PrgJ in *Salmonella* can provide evidence for the substrate switching mechanism [82]. In addition to the inner-rod protein PrgJ, researchers have also identified a homologue of ruler protein, InvJ, in *Salmonella* [73]. By varying the concentration of InvJ as well as PrgJ, one can find whether the substrate switching or the ruler mechanism is operational for that species. It is clear that a combination of quantitative models and experiments will be crucial for building a complete understanding of the assembly and regulation of these massive extracellular structures.

Chapter 4

Concluding remarks

Bacterial secretion systems are intriguing protein export pathways that play an important role in a wide range of cellular functions. The T3SS is a major virulence factor that is present in most gram-negative bacteria and has gained much-deserved attention in the past decade. Advanced experimental techniques have uncovered many crucial aspects of the secretion pathway and researchers have characterized its role in the assembly of essential structures such as the flagellum and the injectisome. Despite the enormous progress made towards understanding the inner workings of this secretion system, there are numerous gaps that are still present. Conflicting models about the length regulation mechanisms in the T3SS have been proposed, and while there is qualitative experimental evidence, there is a lack of a rigorous quantitative framework to characterize these mechanisms. In this work we constructed mathematical models for the two most popular length control mechanisms: the substrate switching and the ruler mechanisms. Not only do our models make quantitative assertions about the substrate switching and the ruler mechanisms, but they also make experimentally testable predictions that will help further validate these mechanisms.

Analysis of the substrate switching model reveals that there is a quadratic relationship between the variance and the average needle lengths. Comparing the results of our model with the available experimental data suggests that *Salmonella* requires about six inner-rod proteins for growing its needles. This prediction is consistent with results of recent mass spectroscopy studies [44, 82]. Our model for the ruler mechanism suggests that an error-prone ruler, as opposed to an exact ruler, can explain the available experimental data. Furthermore, an error-prone ruler allows for a more robust control over lengths by co-varying parameters such as ruler and needle protein production rates. Note that the superior length control obtained by the ruler mechanism comes with an energy cost of producing and exporting ruler proteins, and while this cost is low in the case of the substrate switching mechanism, it is not an efficient mechanism for constructing longer structures since it results in a quadratic increase in the variance. Thus, along with explaining the available experimental data, our models also shed light on the evolutionary trade-offs that might have caused bacterial species to adopt different mechanisms for length regulation.

More work needs to be done to fully characterize the length control mechanisms. In the preceding chapters we discussed specific experiments that will help gain further insights about the length control mechanisms. Our work on the ruler mechanism does not include the analytical calculation for the error-prone ruler and it will be useful to compare the analytical predictions with the results of our stochastic simulations. Experiments in *Salmonella* show evidence for both the substrate switching as well as the ruler mechanism. Perhaps one could formulate a hybrid model that captures the favorable aspects of both these mechanisms. It is clear that a combination of quantitative models and experiments would be crucial for building a complete understanding of the assembly and regulation of T3SS structures.

Bibliography

[1] R. Milo and R. Phillips, *Cell biology by the numbers*, 2015. [Online]. Available: http://book.bionumbers.org/how-many-proteins-are-in-a-cell/

[2] E. R. Green and J. Mecsas, "Bacterial secretion systems - an overview," *Microbiol Spectr*, vol. 4, no. 1, pp. 10.1128/microbiolspec.VMBF–0012–2015, Feb 2016, 26999395[pmid]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4804464/

[3] P. Natale, T. Brüser, and A. J. Driessen, "Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1778, no. 9, pp. 1735 – 1756, 2008, structural proteomics of the cell envelope of Gram-negative bacteria. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005273607002775

[4] S. Thomas, I. B. Holland, and L. Schmitt, "The type 1 secretion pathway — the hemolysin system and beyond," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1843, no. 8, pp. 1629 – 1641, 2014, protein trafficking and secretion in bacteria. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016748891300339X

[5] P. Delepelaire, "Type I secretion in gram-negative bacteria," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1694, no. 1, pp.

149 – 161, 2004, protein Export/Secretion in Bacteria. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167488904001168

[6] K. V. Korotkov, M. Sandkvist, and W. G. J. Hol, "The type II secretion system: biogenesis, molecular architecture and mechanism," *Nature Reviews Microbiology*, vol. 10, pp. 336 EP –, Apr 2012, review Article. [Online]. Available: http://dx.doi.org/10.1038/nrmicro2762

[7] R. Voulhoux, G. Ball, B. Ize, M. L. Vasil, A. Lazdunski, L.-F. Wu, and A. Filloux, "Involvement of the twin-arginine translocation system in protein secretion via the type II pathway," *The EMBO Journal*, vol. 20, no. 23, pp. 6735–6741, 2001. [Online]. Available: http://emboj.embopress.org/content/20/23/6735

[8] G. R. Cornelis, "The type III secretion injectisome," *Nature Reviews Microbiology*, vol. 4, pp. 8117–825, 2006. [Online]. Available: http://dx.doi.org/10.1038/nrmicro1526

[9] R. M. Macnab, "How bacteria assemble flagella," *Annual Review of Microbiology*, vol. 57, no. 1, pp. 77–100, 2003, pMID: 12730325. [Online]. Available: https://doi.org/10.1146/annurev.micro.57.030502.090832

[10] U. Gophna, E. Z. Ron, and D. Graur, "Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events," *Gene*, vol. 312, pp. 151 – 163, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378111903006127

[11] E. Cascales and P. J. Christie, "The versatile bacterial type IV secretion systems," *Nature Reviews Microbiology*, vol. 1, pp. 137 EP –, Nov 2003, review Article. [Online]. Available: http://dx.doi.org/10.1038/nrmicro753

[12] K. Wallden, A. Rivera-Calzada, and G. Waksman, "Type IV secretion

systems: versatility and diversity in function," *Cell Microbiol*, vol. 12, no. 9, pp. 1203–1212, Sep 2010, 20642798[pmid]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3070162/

[13] J. Pohlner, R. Halter, K. Beyreuther, and T. F. Meyer, "Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease," *Nature*, vol. 325, pp. 458 EP –, Jan 1987. [Online]. Available: http://dx.doi.org/10.1038/325458a0

[14] D. L. Leyton, A. E. Rossiter, and I. R. Henderson, "From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis," *Nature Reviews Microbiology*, vol. 10, pp. 213 EP –, Feb 2012, review Article. [Online]. Available: http://dx.doi.org/10.1038/nrmicro2733

[15] A. B. Russell, S. B. Peterson, and J. D. Mougous, "Type VI secretion system effectors: poisons with a purpose," *Nature Reviews Microbiology*, vol. 12, pp. 137 EP –, Jan 2014, review Article. [Online]. Available: http://dx.doi.org/10.1038/nrmicro3185

[16] S. Pukatzki, A. T. Ma, A. T. Revel, D. Sturtevant, and J. J. Mekalanos, "Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin," *Proceedings of the National Academy of Sciences*, vol. 104, no. 39, pp. 15 508–15 513, 2007. [Online]. Available: http://www.pnas.org/content/104/39/15508

[17] J. L. Telford, M. A. Barocchi, I. Margarit, R. Rappuoli, and G. Grandi, "Pili in gram-positive pathogens," *Nature Reviews Microbiology*, vol. 4, pp. 509 EP –, Jul 2006, review Article. [Online]. Available: http://dx.doi.org/10.1038/nrmicro1443

[18] S. K. Mazmanian, G. Liu, H. Ton-That, and O. Schneewind, "*Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall," *Science*, vol. 285, no. 5428, pp. 760–763, 1999. [Online]. Available: http://science.sciencemag.org/content/285/5428/760

[19] R. Simeone, D. Bottai, and R. Brosch, "Esx/type VII secretion systems and their role in host–pathogen interaction," *Current Opinion in Microbiology*, vol. 12, no. 1, pp. 4 – 10, 2009, host–Microbe Interactions: Bacteria. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1369527408001872

[20] E. N. Houben, K. V. Korotkov, and W. Bitter, "Take five — type VII secretion systems of mycobacteria," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1843, no. 8, pp. 1707 – 1716, 2014, protein trafficking and secretion in bacteria. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167488913003820

[21] E. N. G. Houben, J. Bestebroer, R. Ummels, L. Wilson, S. R. Piersma, C. R. Jiménez, T. H. M. Ottenhoff, J. Luirink, and W. Bitter, "Composition of the type VII secretion system membrane complex," *Molecular Microbiology*, vol. 86, no. 2, pp. 472–484. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2012.08206.x

[22] D. Büttner, "Protein export according to schedule: Architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria," *Microbiol Mol Biol Rev*, vol. 76, no. 2, pp. 262–310, Jun 2012, 05017-11[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3372255/

[23] J. van der Heijden and B. B. Finlay, "Type III effector-mediated processes in *Salmonella* infection," *Future Microbiology*, vol. 7, no. 6, pp. 685–703, 2012, pMID: 22702524. [Online]. Available: https://doi.org/10.2217/fmb.12.49

[24] A. S. Santos and B. B. Finlay, "Bringing down the host: enteropathogenic and enterohaemorrhagic *Escherichia coli* effector-mediated subversion of host innate immune pathways," *Cellular Microbiology*, vol. 17, no. 3, pp. 318–332. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cmi.12412

[25] J. B. Bliska, X. Wang, G. I. Viboud, and I. E. Brodsky, "Modulation of innate immune responses by *Yersinia* type III secretion system translocators and effectors," *Cellular Microbiology*, vol. 15, no. 10, pp. 1622–1631. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cmi.12164

[26] W. Deng, N. C. Marshall, J. L. Rowland, J. M. McCoy, L. J. Worrall, A. S. Santos, N. C. J. Strynadka, and B. B. Finlay, "Assembly, structure, function and regulation of type III secretion systems," *Nature Reviews Microbiology*, vol. 15, pp. 323 EP –, Apr 2017, review Article. [Online]. Available: http://dx.doi.org/10.1038/nrmicro.2017.20

[27] F. Gijsegem, C. Gough, C. Zischek, E. Niqueux, M. Arlat, S. Genin, P. Barberis, S. German, P. Castello, and C. Boucher, "The hrp gene locus of *Pseudomonas solanacearum*, which controls the production of a type III secretion system, encodes eight proteins related to components of the bacterial flagellar biogenesis complex," *Molecular Microbiology*, vol. 15, no. 6, pp. 1095–1114. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.1995.tb02284.x

[28] K. A. Fields, G. V. Plano, and S. C. Straley, "A low-Ca2+ response (LCR) secretion *(ysc)* locus lies within the *lcrB* region of the LCR plasmid in *Yersinia pestis*." *J Bacteriol*, vol. 176, no. 3, pp. 569–579, Feb 1994, 8300512[pmid]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC205092/

[29] P. W. G. R. C. Sophie Woestyn, Abdelmounaïm Allaoui, "YscN, the putative energizer of the *Yersinia* Yop secretion machinery," *J Bacteriol*, vol. 176, no. 6, pp. 1561–1569, 1994. [Online]. Available: http://jb.asm.org/content/176/6/1561

[30] C. J. Hueck, "Type III protein secretion systems in bacterial pathogens of animals and plants," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 2, pp. 379–433, 1998. [Online]. Available: http://mmbr.asm.org/content/62/2/379.full.pdf

[31] O. Schraidt, M. D. Lefebre, M. J. Brunner, W. H. Schmied, A. Schmidt, J. Radics, K. Mechtler, J. E. Galán, and T. C. Marlovits, "Topology and organization of the *Salmonella typhimurium* type III secretion needle complex components," *PLoS Pathogens*, vol. 6, no. 4, pp. 1–12, 04 2010. [Online]. Available: https://doi.org/10.1371/journal.ppat.1000824

[32] O. Schraidt and T. C. Marlovits, "Three-dimensional model of *Salmonella*'s needle complex at subnanometer resolution," *Science*, vol. 331, no. 6021, pp. 1192–1195, 2011. [Online]. Available: http://science.sciencemag.org/content/331/6021/1192

[33] B. J. Burkinshaw and N. C. Strynadka, "Assembly and structure of the T3SS," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1843, no. 8, pp. 1649 – 1663, 2014, protein trafficking and secretion in bacteria. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167488914000470

[34] L. J. Worrall, C. Hong, M. Vuckovic, W. Deng, J. R. C. Bergeron, D. D. Majewski, R. K. Huang, T. Spreter, B. B. Finlay, Z. Yu, and N. C. J. Strynadka, "Near-atomic-resolution cryo-EM analysis of the *Salmonella* T3S injectisome basal body," *Nature*, vol. 540, pp. 597 EP –, Dec 2016. [Online]. Available: http://dx.doi.org/10.1038/nature20576

[35] J. L. Hodgkinson, A. Horsley, D. Stabat, M. Simon, S. Johnson, P. C. A. da Fonseca, E. P. Morris, J. S. Wall, S. M. Lea, and A. J. Blocker, "Three-dimensional reconstruction of the *Shigella* T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout," *Nature Structural &Amp; Molecular Biology*, vol. 16, pp. 477 EP –, Apr 2009, article. [Online]. Available: http://dx.doi.org/10.1038/nsmb.1599

[36] J. Kowal, M. Chami, P. Ringler, S. Müller, M. Kudryashev, D. Castaño-Díez, M. Amstutz, G. Cornelis, H. Stahlberg, and A. Engel, "Structure of the dodecameric *Yersinia enterocolitica* secretin YscC and its trypsin-resistant

core," *Structure*, vol. 21, no. 12, pp. 2152–2161, Dec 2013. [Online]. Available: http://dx.doi.org/10.1016/j.str.2013.09.012

[37] J. E. Deane, S. C. Graham, E. P. Mitchell, D. Flot, S. Johnson, and S. M. Lea, "Crystal structure of Spa40, the specificity switch for the *Shigella flexneri* type III secretion system," *Mol Microbiol*, vol. 69, no. 1, pp. 267–276, Jul 2008, 18485071[pmid]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2615192/

[38] R. Zarivach, W. Deng, M. Vuckovic, H. B. Felise, H. V. Nguyen, S. I. Miller, B. B. Finlay, and N. C. J. Strynadka, "Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS," *Nature*, vol. 453, pp. 124 EP –, May 2008. [Online]. Available: http://dx.doi.org/10.1038/nature06832

[39] U. Wiesand, I. Sorg, M. Amstutz, S. Wagner, J. van den Heuvel, T. Lührs, G. R. Cornelis, and D. W. Heinz, "Structure of the type III secretion recognition protein YscU from *Yersinia enterocolitica*," *Journal of Molecular Biology*, vol. 385, no. 3, pp. 854 – 866, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022283608013119

[40] R. Zarivach, M. Vuckovic, W. Deng, B. B. Finlay, and N. C. J. Strynadka, "Structural analysis of a prototypical ATPase from the type III secretion system," *Nature Structural &Amp; Molecular Biology*, vol. 14, pp. 131 EP –, Jan 2007, article. [Online]. Available: http://dx.doi.org/10.1038/nsmb1196

[41] K. Imada, T. Minamino, Y. Uchida, M. Kinoshita, and K. Namba, "Insight into the flagella type III export revealed by the complex structure of the type III atpase and its regulator," *Proceedings of the National Academy of Sciences*, vol. 113, no. 13, pp. 3633–3638, 2016. [Online]. Available: http://www.pnas.org/content/113/13/3633

[42] L. Claret, S. R. Calder, M. Higgins, and C. Hughes, "Oligomerization and

activation of the FliI ATPase central to bacterial flagellum assembly," *Mol Microbiol*, vol. 48, no. 5, pp. 1349–1355, Jun 2003, 12787361[pmid]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2528289/

[43] S. A. Müller, C. Pozidis, R. Stone, C. Meesters, M. Chami, A. Engel, A. Economou, and H. Stahlberg, "Double hexameric ring assembly of the type III protein translocase ATPase HrcN," *Molecular Microbiology*, vol. 61, no. 1, pp. 119–125. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2006.05219.x

[44] S. Zilkenat, M. Franz-Wachtel, Y.-D. Stierhof, J. E. Galán, B. Macek, and S. Wagner, "Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach," *Mol Cell Proteomics*, vol. 15, no. 5, pp. 1598–1609, May 2016, m115.056598[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4858942/

[45] T. G. Kimbrough and S. I. Miller, "Contribution of *Salmonella typhimurium* type III secretion components to needle complex formation," *Proceedings of the National Academy of Sciences*, vol. 97, no. 20, pp. 11 008–11 013, 2000. [Online]. Available: http://www.pnas.org/content/97/20/11008

[46] E. Hoiczyk and G. Blobel, "Polymerization of a single protein of the pathogen *Yersinia enterocolitica* into needles punctures eukaryotic cells," *Proc Natl Acad Sci U S A*, vol. 98, no. 8, pp. 4669–4674, Apr 2001, 071065798[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC31892/

[47] A. Loquet, N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker, and A. Lange, "Atomic model of the type III secretion system needle," *Nature*, vol. 486, no. 7402, pp. 276–279, May 2012, 22699623[pmid]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3598588/

[48] J.-P. Demers, N. G. Sgourakis, R. Gupta, A. Loquet, K. Giller, D. Riedel, B. Laube, M. Kolbe, D. Baker, S. Becker, and A. Lange, "The common structural architecture of *Shigella flexneri* and *Salmonella typhimurium* type three secretion needles," *PLoS Pathogens*, vol. 9, no. 3, pp. 1–11, 03 2013. [Online]. Available: https://doi.org/10.1371/journal.ppat.1003245

[49] C. A. Mueller, P. Broz, S. A. Müller, P. Ringler, F. Erne-Brand, I. Sorg, M. Kuhn, A. Engel, and G. R. Cornelis, "The V-antigen of *Yersinia* forms a distinct structure at the tip of injectisome needles," *Science*, vol. 310, no. 5748, pp. 674–676, 2005. [Online]. Available: http://science.sciencemag.org/content/310/5748/674

[50] C. R. Epler, N. E. Dickenson, E. Bullitt, and W. L. Picking, "Ultrastructural analysis of IpaD at the tip of the nascent MxiH type III secretion apparatus of *Shigella flexneri*," *Journal of Molecular Biology*, vol. 420, no. 1, pp. 29–39, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022283612002999

[51] A. Blocker, P. Gounon, E. Larquet, K. Niebuhr, V. Cabiaux, C. Parsot, and P. Sansonetti, "The tripartite type III secreton of *Shigella flexneri* inserts IpaB and IpaB into host membranes," *The Journal of Cell Biology*, vol. 147, no. 3, pp. 683–693, 1999. [Online]. Available: http://jcb.rupress.org/content/147/3/683

[52] W. L. Picking, H. Nishioka, P. D. Hearn, M. A. Baxter, A. T. Harrington, A. Blocker, and W. D. Picking, "IpaD of *Shigella flexneri* is independently required for regulation of ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes," *Infect Immun*, vol. 73, no. 3, pp. 1432–1440, Mar 2005, 1410-04[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1064949/

[53] M. Espina, A. J. Olive, R. Kenjale, D. S. Moore, S. F. Ausar, R. W. Kaminski, E. V. Oaks, C. R. Middaugh, W. D. Picking, and W. L. Picking, "IpaD localizes

to the tip of the type III secretion system needle of *Shigella flexneri*," *Infect Immun*, vol. 74, no. 8, pp. 4391–4400, Aug 2006, 0440-06[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1539624/

[54] Q. Jin and S.-Y. He, "Role of the Hrp pilus in type III protein secretion in *Pseudomonas syringae*," *Science*, vol. 294, no. 5551, pp. 2556–2558, 2001. [Online]. Available: http://science.sciencemag.org/content/294/5551/2556

[55] S. Knutton, I. Rosenshine, M. J. Pallen, I. Nisan, B. C. Neves, C. Bain, C. Wolff, G. Dougan, and G. Frankel, "A novel EspA-associated surface organelle of enteropathogenic *Escherichia coli* involved in protein translocation into epithelial cells," *The EMBO Journal*, vol. 17, no. 8, pp. 2166–2176, 1998. [Online]. Available: http://emboj.embopress.org/content/17/8/2166

[56] C. K. Yip, B. B. Finlay, and N. C. J. Strynadka, "Structural characterization of a type iii secretion system filament protein in complex with its chaperone," *Nature Structural &Amp; Molecular Biology*, vol. 12, pp. 75 EP –, Dec 2004, article. [Online]. Available: http://dx.doi.org/10.1038/nsmb879

[57] S. S. Abby and E. P. C. Rocha, "The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems," *PLoS Genetics*, vol. 8, no. 9, pp. 1–15, 09 2012. [Online]. Available: https://doi.org/10.1371/journal.pgen.1002983

[58] F. Makino, D. Shen, N. Kajimura, A. Kawamoto, P. Pissaridou, H. Oswin, M. Pain, I. Murillo, K. Namba, and A. J. Blocker, "The architecture of the cytoplasmic region of type iii secretion systems," *Scientific Reports*, vol. 6, pp. 33 341 EP –, Sep 2016, article. [Online]. Available: http://dx.doi.org/10.1038/srep33341

[59] S. Kojima and D. F. Blair, "The bacterial flagellar motor: Structure and

function of a complex molecular machine," ser. International Review of Cytology. Academic Press, 2004, vol. 233, pp. 93 – 134. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0074769604330032

[60] Y. Sowa and R. M. Berry, "Bacterial flagellar motor," *Quarterly Reviews of Biophysics*, vol. 41, no. 2, p. 103–132, 2008.

[61] N. R. Francis, G. E. Sosinsky, D. Thomas, and D. J. DeRosier, "Isolation, characterization and structure of bacterial flagellar motors containing the switch complex," *Journal of Molecular Biology*, vol. 235, no. 4, pp. 1261 – 1270, 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022283684710795

[62] T. Hirano, T. Minamino, K. Namba, and R. M. Macnab, "Substrate specificity classes and the recognition signal for salmonella type iii flagellar export," *J Bacteriol*, vol. 185, no. 8, pp. 2485–2492, Apr 2003, 1332[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC152621/

[63] F. Cordes, K. Komoriya, E. Larquet, S. Yang, E. H. Egelman, A. Blocker, and S. M. Lea, "Helical structure of the needle of the type III secretion system of *Shigella flexneri*," *The Journal of Biological Chemistry*, vol. 278, no. 19.

[64] A. Gauthier, J. L. Puente, and B. B. Finlay, "Secretin of the enteropathogenic *Escherichia coli* type III secretion system requires components of the type III apparatus for assembly and localization," *Infect Immun*, vol. 71, no. 6, pp. 3310–3319, Jun 2003, 1341[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC155723/

[65] S. Wagner, L. Königsmaier, M. Lara-Tejero, M. Lefebre, T. C. Marlovits, and J. E. Galán, "Organization and coordinated assembly of the type III secretion export apparatus," *Proceedings of the National Academy of Sciences*, vol. 107, no. 41, pp. 17 745–17 750, 2010. [Online]. Available: http://www.pnas.org/content/107/41/17745

[66] S. Shibata, N. Takahashi, F. F. V. Chevance, J. E. Karlinsey, K. T. Hughes, and S. Aizawa, "FliK regulates flagellar hook length as an internal ruler," *Molecular Microbiology*, vol. 64, no. 5, pp. 1404–1415. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2007.05750.x

[67] S. Makishima, K. Komoriya, S. Yamaguchi, and S.-I. Aizawa, "Length of the flagellar hook and the capacity of the type III export apparatus," *Science*, vol. 291, no. 5512, pp. 2411–2413, 2001. [Online]. Available: http://science.sciencemag.org/content/291/5512/2411

[68] B. Hu, D. R. Morado, W. Margolin, J. R. Rohde, O. Arizmendi, W. L. Picking, W. D. Picking, and J. Liu, "Visualization of the type III secretion sorting platform of *Shigella flexneri*," *Proceedings of the National Academy of Sciences*, vol. 112, no. 4, pp. 1047–1052, 2015. [Online]. Available: http://www.pnas.org/content/112/4/1047

[69] A. Kawamoto, Y. V. Morimoto, T. Miyata, T. Minamino, K. T. Hughes, T. Kato, and K. Namba, "Common and distinct structural features of *Salmonella* injectisome and flagellar basal body," *Scientific Reports*, vol. 3, pp. 3369 EP –, Nov 2013, article. [Online]. Available: http://dx.doi.org/10.1038/srep03369

[70] M. Kudryashev, M. Stenta, S. Schmelz, M. Amstutz, U. Wiesand, D. Castaño Díez, M. T. Degiacomi, S. Münnich, C. K. Bleck, J. Kowal, A. Diepold, D. W. Heinz, M. Dal Peraro, G. R. Cornelis, and H. Stahlberg, "*In situ* structural analysis of the *Yersinia enterocolitica* injectisome," *eLife*, vol. 2, p. e00792, jul 2013. [Online]. Available: https://doi.org/10.7554/eLife.00792

[71] A. Nans, M. Kudryashev, H. R. Saibil, and R. D. Hayward, "Structure of a bacterial type III secretion system in contact with a host membrane *in situ*," *Nature*

Communications, vol. 6, pp. 10 114 EP –, Dec 2015, article. [Online]. Available: http://dx.doi.org/10.1038/ncomms10114

[72] L. Journet, C. Agrain, P. Broz, and G. R. Cornelis, "The needle length of bacterial injectisomes is determined by a molecular ruler," *Science*, vol. 302, no. 5651, pp. 1757–1760, 2003. [Online]. Available: http://science.sciencemag.org/content/302/5651/1757

[73] D. H. Wee and K. T. Hughes, "Molecular ruler determines needle length for the *Salmonella* SPI-1 injectisome," *Proceedings of the National Academy of Sciences*, vol. 112, no. 13, pp. 4098–4103, 2015. [Online]. Available: http://www.pnas.org/content/112/13/4098

[74] S. Wagner, M. Stenta, L. C. Metzger, M. Dal Peraro, and G. R. Cornelis, "Length control of the injectisome needle requires only one molecule of Yop secretion protein p (YscP)," *Proceedings of the National Academy of Sciences*, vol. 107, no. 31, pp. 13 860–13 865, 2010. [Online]. Available: http://www.pnas.org/content/107/31/13860

[75] T. Minamino, Y. Saijo-Hamano, Y. Furukawa, B. González-Pedrajo, R. M. Macnab, and K. Namba, "Domain organization and function of *Salmonella* FliK, a flagellar hook-length control protein," *Journal of Molecular Biology*, vol. 341, no. 2, pp. 491 – 502, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S002228360400676X

[76] M. Erhardt, H. M. Singer, D. H. Wee, J. P. Keener, and K. T. Hughes, "An infrequent molecular ruler controls flagellar hook length in *Salmonella enterica*," *The EMBO Journal*, vol. 30, no. 14, pp. 2948–2961, 2011. [Online]. Available: http://emboj.embopress.org/content/30/14/2948

[77] T. C. Marlovits, T. Kubori, M. Lara-Tejero, D. Thomas, V. M. Unger, and J. E. Galán, "Assembly of the inner rod determines needle length in the type III

secretion injectisome," *Nature*, vol. 441, pp. 637 EP –, Jun 2006. [Online]. Available: http://dx.doi.org/10.1038/nature04822

[78] T. C. Marlovits, T. Kubori, A. Sukhan, D. R. Thomas, J. E. Galán, and V. M. Unger, "Structural insights into the assembly of the type III secretion needle complex," *Science*, vol. 306, no. 5698, pp. 1040–1042, 2004. [Online]. Available: http://science.sciencemag.org/content/306/5698/1040

[79] K. Tamano, S.-I. Aizawa, E. Katayama, T. Nonaka, S. Imajoh-Ohmi, A. Kuwae, S. Nagai, and C. Sasakawa, "Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors," *EMBO J*, vol. 19, no. 15, pp. 3876–3887, Aug 2000, cdd394[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC306602/

[80] T. Kubori, A. Sukhan, S.-I. Aizawa, and J. E. Galán, "Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system," *Proceedings of the National Academy of Sciences*, vol. 97, no. 18, pp. 10 225–10 230, 2000. [Online]. Available: http://www.pnas.org/content/97/18/10225

[81] J. Keener, "A molecular ruler mechanism for length control of extended protein structures in bacteria," *Journal of Theoretical Biology*, vol. 263, no. 4, pp. 481 – 489, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022519309005827

[82] M. K. Nariya, J. Israeli, J. J. Shi, and E. J. Deeds, "Mathematical model for length control by the timing of substrate switching in the type III secretion system," *PLoS Computational Biology*, vol. 12, no. 4, pp. 1–14, 04 2016. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1004851

[83] M. K. Nariya, A. Mallela, J. J. Shi, and E. J. Deeds, "Molecular ruler allows for a

tight control of the lengths in the flagellar hook and the type III secretion injectisome," *Unpublished work.*

[84] E. J. Deeds, J. A. Bachman, and W. Fontana, "Optimizing ring assembly reveals the strength of weak interactions," *Proc Natl Acad Sci U S A*, vol. 109, no. 7, pp. 2348–2353, Feb 2012, 201113095[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3289332/

[85] S. Mangan and U. Alon, "Structure and function of the feed-forward loop network motif," *Proceedings of the National Academy of Sciences*, vol. 100, no. 21, pp. 11 980–11 985, 2003. [Online]. Available: http://www.pnas.org/content/100/21/11980

[86] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, Dec 1977. [Online]. Available: https://doi.org/10.1021/j100540a008

[87] V. Danos and C. Laneve, "Formal molecular biology," *Theor. Comput. Sci.*, vol. 325, no. 1, pp. 69–110, Sep. 2004. [Online]. Available: http://dx.doi.org/10.1016/j.tcs.2004.03.065

[88] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine, *Rule-Based Modelling of Cellular Signalling.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 17–41.

[89] R. Suderman and E. J. Deeds, "Machines vs. ensembles: Effective MAPK signaling through heterogeneous sets of protein complexes," *PLoS Computational Biology*, vol. 9, no. 10, pp. 1–11, 10 2013. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1003278

Appendix A

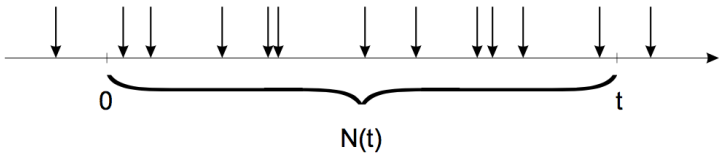
Poisson Processes

Definition

A process is said to be a Poisson process if the number of occurrences in a finite interval of time t obeys a Poisson distribution:

$$P\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \forall t \geq 0$$

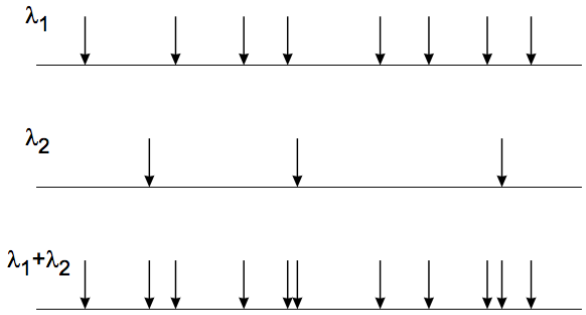
where λ is called the rate of the process.



Important properties

1. The $N(t) = n$ arrivals in the interval of $(0, t)$ are independent and uniformly distributed in the interval.

2. **Superposition:** The superposition of two Poisson process with rates λ_1 and λ_2 is a Poisson process with rate $\lambda_1 + \lambda_2$.



3. **Memoryless:** The remaining time to the next arrival has the same exponential distribution irrespective of the time that has already elapsed since the previous arrival (the same holds also in reversed time, *i.e.* looking backwards). Mathematically this can be written as:

$$P\{X > t + s\} = P\{X > t\}P\{X > s\}\forall t, s \geq 0.$$

Probability density function

As seen in property 1, the n occurrences are independent and identically distributed (IID) random variables. Each one of these n occurrences has a density function, $\mathsf{f}_X(t) = \lambda e^{-\lambda t}$. If S_n denotes the sum of n IID random variables, then the density function for S_n can be calculated by convolving their densities (convolving $\mathsf{f}_X(t)$ with itself for obtaining $\mathsf{f}_{S_2}(t)$, $\mathsf{f}_X(t)$ with density of S_2 for $\mathsf{f}_{S_3}(t)$ and so on). The density function of S_n is known as the *Erlang density*:

$$\mathsf{f}_{S_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}.$$